Plans for a clinical trial of an anti-aging drug *p. 1274*

Inuit metabolism under selection *pp. 1282 & 1343*

Damage control in neural stem cells *p. 1334*

Science

$10
18 SEPTEMBER 2015
sciencemag.org

AAAS

Designed invisibility

A flexible array of nanostructures forms a skin cloak *pp. 1268 & 1310*

CONTENTS

1283 & 1351
Interrogating prostate cancer cells

18 SEPTEMBER 2015 • VOLUME 349 • ISSUE 6254

NEWS

IN BRIEF

IN DEPTH

FEATURE

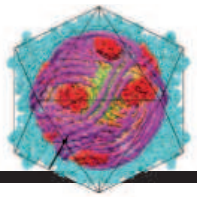
INSIGHTS

PERSPECTIVES

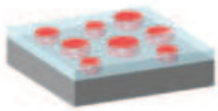
BOOKS *ET AL.*

LETTERS

Published by AAAS



1347
Copying RNA inside a virus



1314
Polarizing a thin film

RESEARCH

IN BRIEF

RESEARCH ARTICLES

REPORTS

DEPARTMENTS

ON THE COVER

Illustration of an ultrathin cloak consisting of nanoantennas (gold blocks) conformally coating an arbitrarily shaped surface, rendering it invisible with red light. The varying dimensions of the nanoantennas correspond to different phase shifts that reroute light according to the changing height of the surface. This nanoscale illustration allows us to see the cloak, which is not visible to the naked eye. In the future, this cloaking mechanism will potentially be scalable for macroscopic objects. See pages 1269 and 1310.
Illustration: Valerie Altounian/Science

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices.

Domestic individual membership and subscription (51 issues): $153 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1282; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request.

Published by AAAS

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Bateman House, 82-88 Hills Road, Cambridge, UK CB2 1LQ

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Services** 202-326-6417 or http://membercentral.aaas.org/discounts

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority of conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and books reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

INFORMATION FOR AUTHORS See pages 678 and 679 of the 6 February 2015 issue or access www.sciencemag.org/about/authors

EDITORIAL

Wildfires burn science capacity

With just over 3 months remaining, it looks like 2015 could be a record-breaking year for wildfires in the United States. So far this year, more than 8.5 million acres have burned and severe fires often happen in October. For the first time, the U.S. Forest Service will spend over 50% of its budget on fire management. Ironically, this is directing resources away from projects and research that would reduce the risk of forest fires. This must be fixed.

Trends show the need for action by Congress. Wildfires are getting more severe; the wildfire season continues to grow longer, increasing in some areas by 7 weeks over the past 35 years. More people are living in the fire-prone wildland/urban interface, raising the risk of damaging wildfires. Vast areas have an unnatural buildup of highly flammable trees and brush due to past overzealous restraining of low-intensity natural fires.

"...Congress should pass the Wildfire Disaster Funding Act."

Not surprisingly, firefighting costs have soared. Indeed, federal appropriations for fire suppression and fire operations zoomed from about $600 million in 1995 to nearly $3 billion in 2014. State fire-suppression expenditures doubled from 1998 to 2014, to $1.6 billion. So far, available appropriations this year for firefighting are $700 million less than actually spent, so non-fire funds have been borrowed to cover these emergency needs. This season, the U.S. Forest Service spent as much as $243 million per week fighting wildfires. Fire-related costs take up over 50% of its budget, up from 16% in 1995. If changes are not made, fire management will consume 67% of its budget by 2025.

Making matters worse is that in the United States, wildfires are the only kind of natural disaster for which a response consumes regular federal agency funding. Wildfire fighting is largely supported by federal appropriations to the Forest Service and the Department of the Interior, along with state, tribal, county, and municipal participation. Paying for firefighting using normal operating funds has vast consequences for both land management and the science that can address fire problems. This science capacity is hurt not only at the Forest Service but also at the U.S. Geological Survey, the U.S. Environmental Protection Agency, the Smithsonian Institution, and others that are funded in the same legislation from Congress. For example, the Forest Service Research and Development arm has had a 33% reduction in staffing from 2001 to 2014.

There is much at stake with this loss of land management and science capacity. Forestlands provide half of our nation's water and sequester about 13% of total U.S. fossil fuel carbon emissions, but projections suggest that forests will become net carbon emitters later this century if steps are not taken to make them more resilient. Science is needed for inventories, for modeling fire behavior and predicting treatment efficacy, and for helping with risk evaluation to improve management decisions.

A 4-year-long effort by all levels of government, including federal, state, tribal, county, and municipal units, developed the National Cohesive Wildland Fire Management Strategy, a plan that uses science and data analysis to support three major goals: restore and maintain fire-adapted landscapes, create fire-adapted communities, and improve wildfire response. The actions include such things as engaging citizens in safer building practices, implementing controlled burns, and removing vegetation to safeguard communities, wildlife, and water resources. All levels of government have agreed to this strategy, but now they must provide resources to make it work. If implemented, it will reduce firefighting costs while enhancing many other benefits to society and nature.

In addition, Congress should pass the Wildfire Disaster Funding Act. Pending in the House and the Senate, the bill would change how the federal government budgets for the suppression of wildfire disasters, making it similar to the way other responses to natural disasters are funded.

– Christopher Topik

Christopher Topik is the director of Restoring America's Forests at The Nature Conservancy North America Region, Arlington, VA. E-mail: ctopik@tnc.org

PHOTOS: (TOP RIGHT) C. TOPIK; (INSET) AP PHOTO/KEVIN WARN

10.1126/science.aad4202

NEWS

> “Its intense aroma had hints of antiseptic smoke, rubber, and smoked fish …”
>
> Tasting notes by Scotland’s **Ardbeg Distillery** on whiskey sent to the International Space Station 4 years ago and returned last year. Pre–space journey, the same whiskey had “hints of cedar, sweet smoke, and aged balsamic vinegar.”

IN BRIEF

Fires disrupt and inspire scientists

A wildfire burns in Montana’s Glacier National Park in July.

Some 45,000 wildfires in the United States so far this year have scorched more than 3.5 million hectares, destroyed hundreds of homes, and killed at least four people. Whereas many scientists bemoan sites ruined by the flames, some also see new research opportunities. In Idaho, the Tepee Springs fire may force the abandonment of a logging experiment that took months to plan, says forest scientist Robert Keefe of the University of Idaho, Moscow. In Montana, blazes have prevented ecologist Tabitha Graves, of the U.S. Geological Survey’s Northern Rocky Mountain Science Center in West Glacier, from reaching cameras that track the growth of huckleberries, a key food for bears. But even if the cameras melt, she says the flames might provide a chance to study postfire recovery. And “when the smoke clears … years like this present great research opportunities,” says fire researcher Max Moritz of the University of California, Berkeley, although finding funding can be tough.

AROUND THE WORLD

Court nixes pesticide approval

SAN FRANCISCO, CALIFORNIA | Concerned about the precarious state of bee populations, a federal court last week ordered the Environmental Protection Agency (EPA) to withdraw its approval of sulfoxaflor, a new type of insecticide, until it has more science at hand. EPA approved sulfoxaflor in May 2013 for many crops, including cotton and citrus. Environmental and beekeeping groups sued, contending that there was inadequate science to rule out impacts on honey bees. EPA argued that it protected bees by approving lower doses and restricting spraying when bees are most active, but the Court of Appeals for the Ninth Circuit ordered the agency to come up with more data on bee impacts. Dow AgroSciences, which manufactures pesticides with sulfoxaflor, said in a statement that it will work with EPA to “complete additional regulatory work” but that it is considering appealing the court’s decision.

Russia absorbs Crimean institutes

SIMFEROPOL, CRIMEA | After months in limbo, scientists in a Black Sea region annexed by Russia could see their prospects improve. Last week, the Russian government announced that the Russian Academy of Sciences (RAS) will absorb the cream of Crimea’s research institutes. After Russia seized control of the Crimean Peninsula in March 2014, the United States, the European Union, and other countries sanctioned Russian and Crimean

Crimean Astrophysical Observatory.

PHOTOS (LEFT TO RIGHT): STRINGER/REUTERS/NEWSCOM; © KONSTANTIN LABUNSKIY/ALAMY STOCK PHOTO

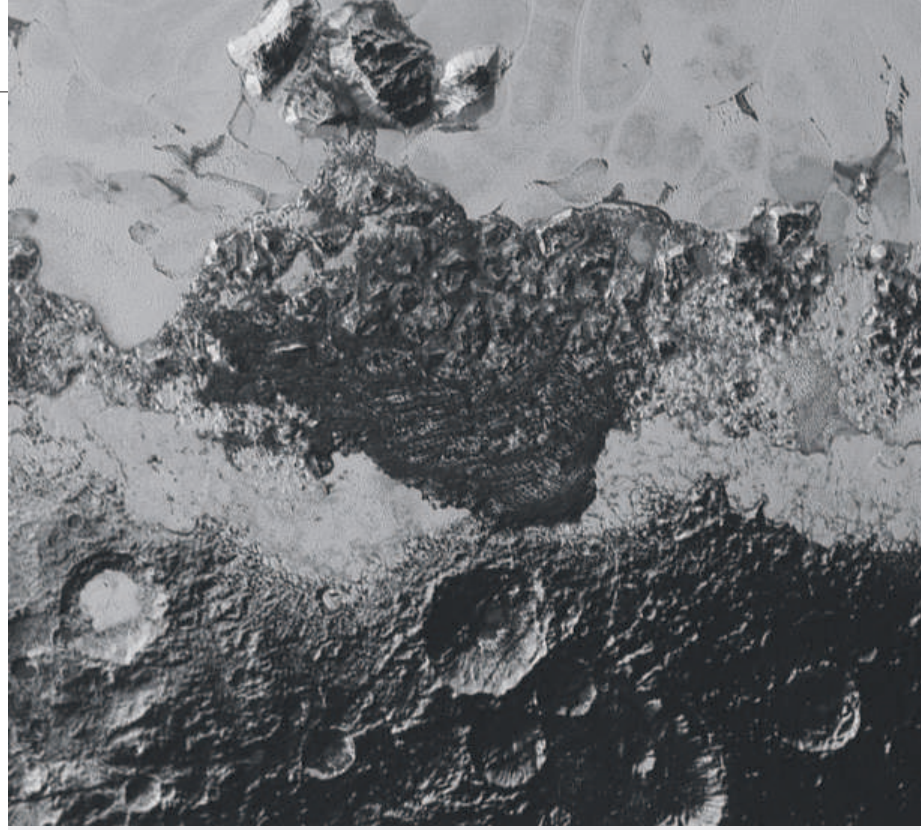
entities that facilitated Russia's takeover. Most scientific cooperation with Crimea stopped, and key institutes were placed under the control of local administrators lacking scientific expertise. But some scientists in Crimea who welcomed annexation note that their funding has improved because they can compete for Russian grants. Now, several research centers—including the Crimean Astrophysical Observatory in Nauchny and the Institute of Archaeology in Simferopol—will become part of RAS. "It means for us that our research will be evaluated now by real scientists. We fought for this status and we won," says Valentina Mordvintseva of the Institute of Archaeology.

Bioethics group OKs gene editing

MANCHESTER, U.K. | Genetic editing of human embryos "has tremendous value" to help solve important scientific questions, and should proceed despite ethical concerns, the Hinxton Group, an influential international bioethics consortium, said on 9 September. New techniques that allow researchers to precisely edit genes in living cells have become powerful tools but have raised questions about the ethics of genetically altering humans in ways that could be passed on to future generations; for example, in April, Chinese scientists published the first paper describing the use of a genome editing technique called CRISPR/Cas9 in human embryos. The Hinxton Group called for more discussion and careful policies to govern research using gene editing in embryos, but concluded that the insights such research could provide into early human development and disease are ethically justifiable. They also concluded that use of the technologies for reproduction is premature.
http://scim.ag/Hinxtongeneedit

Safety probe at DOD labs

ATLANTA | The Centers for Disease Control and Prevention (CDC) is investigating lapses in handling dangerous pathogens at four Department of Defense (DOD) biodefense labs. The probe follows the discovery in May that over the past decade, DOD's Dugway lab in Utah accidentally shipped live anthrax to scores of other labs. In August, CDC inspectors found recordkeeping problems with samples of plague bacteria and several viruses at a DOD lab in Edgewood, Maryland. That triggered the Army on 2 September to order a safety review at all nine DOD labs working with so-called select agents, a list of high-risk pathogens. Army officials also suspended select agent work at Dugway, Edgewood, and two other Maryland labs pending the outcome of safety reviews.



Dark ridges at the bottom of bright Sputnik Planum may be dunes; old, cratered terrain is visible below that.

PHOTO: NASA/JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY/SOUTHWEST RESEARCH INSTITUTE

Pluto and Charon's complicated faces

NASA has released more images of Pluto and its moon Charon, taken by its New Horizons spacecraft during its July flyby. The new images of Pluto reveal more of the dwarf planet's complex surface, including a mishmash of features near a flat, icy plain informally named Sputnik Planum (one lobe of Tombaugh Regio, also known as Pluto's bright "heart"). Chaotic, jumbled mountains at one edge of the plain may be blocks of water ice floating within a denser, softer deposit of frozen nitrogen, says Jeff Moore, a New Horizons team member at NASA's Ames Research Center in Mountain View, California. The icy plain also appears to contain dark, windswept dunes, even though Pluto's atmosphere is thought to be too thin for significant wind. "The surface of Pluto is every bit as complex as that of Mars," Moore said in a statement. New images of Charon, meanwhile, show the moon's dark, frigid poles—just tens of degrees warmer than absolute zero—in greater detail, as well as a reddish blemish that appears to be an impact basin.

Intel exits school science contest

SANTA CLARA, CALIFORNIA | The Intel Science Talent Search, one of the nation's most prestigious competitions for science-savvy high school students in the United States, is losing its title sponsor. Intel has announced that it will no longer sponsor the program, and the nonprofit that runs the competition, the Society for Science & the Public in Washington, D.C., is looking for a new sponsor to pick up the $6 million annual tab starting in 2017. The program, meant to "inspire innovators of tomorrow," targets science, math, engineering, and technology students in their last year of high school. It has drawn in thousands of hopeful applicants since it began in 1942. The winners receive prize money ranging from $35,000 to $150,000.

A science-friendlier leader?

SYDNEY, AUSTRALIA | Australian scientists are hopeful of a better hearing from their government after conservative Prime Minister Tony Abbott was unceremoniously dumped from his post Monday evening by members of his own Liberal party. In office, Abbott oversaw deep cuts to science and research funding, shuttered the independent Climate Commission, and failed to appoint a science minister for more than a year. Abbott's replacement,

Malcolm Turnbull, has shown support for the science of climate change, and had previously supported the Labor Government's Carbon Pollution Reduction Scheme during his tenure as opposition leader. Abbott, who dismissed climate change as "absolute crap," ousted Turnbull as the leader of the Liberal party in 2009. But Turnbull now says he will make no immediate changes to Abbott's climate policy; however, he has promised to put innovation and technology at the center of his government.

NEWSMAKERS

Physicist fraud case dropped

Federal prosecutors filed a motion on 11 September in Philadelphia, Pennsylvania, to drop a case against Temple University physicist **Xiaoxing Xi**, who had been accused of helping Chinese organizations illegally obtain U.S. technology. In a 14 May indictment, the government alleged that Xi, an expert on thin-film materials, schemed to pass information about a device known as a Pocket Heater—a proprietary U.S. technology used to make magnesium diboride superconducting thin films—to Chinese entities to help them become leaders in the field of superconductivity. Citing four email messages between Xi and colleagues in China, federal investigators charged him with four counts of wire fraud. In June, Xi pleaded not guilty to the charges; his lawyer argued that the email exchanges concern "routine academic collaboration" and discussed technologies that "were not restricted in any way." http://scim.ag/XiaoxingXi

Three Q's

The United States lacks a world-class education system despite conducting more education research than any other country. **Linda Darling-Hammond**, professor emeritus of education at Stanford University in Palo Alto, California, hopes her new Learning Policy Institute, unveiled this month, will put more of that research to use by making it understandable and accessible to policymakers. http://scim.ag/HammondLPI

Q: Why create another think tank?
A: Our core concerns involve a set of issues you might call 21st century learning. But what we know about how to develop this kind of learning is very, very far away from policy.

Q: What do you mean by that phrase?
A: Our scientific knowledge is expanding so rapidly, you don't want to have kids just memorize a bunch of facts that will soon be out of date. Students need to acquire and find knowledge and make sense of it themselves. That's a radically different kind of teaching than what most schools now offer. And transforming teaching may mean transforming school organizations.

Q: How will your institute stand out from the pack?
A: Research can't solve deep ideological divides. But we will put more boots on the ground to make sure that the research is getting translated and available at the moment when it is needed.

An artist's rendition depicts standing stones at a possible Neolithic ritual site.

'Superhenge' found near Stonehenge?

Scientists may have found the remains of a huge prehistoric stone monument—buried under a grassy bank about 3 kilometers from Stonehenge. Using remote sensing and geophysical imaging technology, a team of researchers with the Stonehenge Hidden Landscapes Project has been creating an underground map of the region since July 2010. Last week at the British Science Festival in Bradford, U.K., the team announced that they had discovered evidence of as many as 90 standing stones arranged in the shape of a C around a natural depression. The monument is thought to be a Neolithic ritual site, dating back to more than 4500 years ago. At about five times the size of Stonehenge, the site would be the largest surviving stone monument found in the United Kingdom, and possibly in Europe, the researchers say. But without excavation, what the scientists are calling standing stones could be a number of things, including rows of large pits or natural features, according to Mike Pitts, a megaliths expert and editor of *British Archaeology*.

BY THE NUMBERS

57%

Klebsiella pneumoniae isolates in India that were resitant to antibiotics of last resort in 2014, up from 29% in 2008. In most of Europe, the number is below 5% (Center for Disease Dynamics, Economics, & Policy).

$1 billion

Amount of money that bats save corn farmers around the world each year by feeding on corn earworms, thereby increasing crop yields by 1.4% (*Proceedings of the National Academy of Sciences*).

73%

Amount of Americans who can distinguish the definition of astrology from astronomy (Pew Research Center).

Antiabortion protestors call for ending federal funding of Planned Parenthood in Washington, D.C., earlier this year.

SCIENCE POLICY

In Wisconsin, an early clash over fetal tissue

Planned Parenthood controversy spurs state bill that would criminalize research

By **Kelly Servick**

A conflict is escalating over U.S. researchers' use of human fetal tissue. Legislators in Wisconsin last week advanced a bill that would make it a felony for scientists working in the state to conduct studies using tissue or cells obtained from recently aborted fetuses. The measure, approved by a committee of the Wisconsin State Assembly, has drawn opposition from universities and research groups, who say it will stifle important disease studies. The bill is likely just the first of many similar state-level efforts, science policy observers predict.

"We're expecting to see a wave of legislation in 2016," says Elizabeth Nash, senior state issues associate at the Guttmacher Institute, a nonprofit reproductive health organization in Washington, D.C. "In a number of places, the chance that these [proposals] become state law is very high."

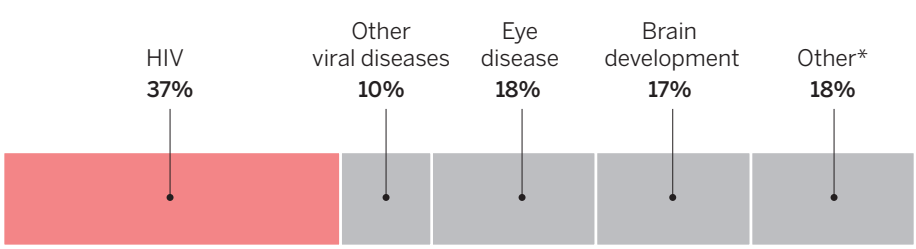
The surge in opposition to the use of tissue obtained through abortions is largely a reaction to hidden-camera footage released earlier this year by the Center for Medical Progress, an antiabortion group based in Irvine, California. The videos show Planned Parenthood officials discussing how the group fulfills requests from researchers for samples from aborted fetuses. Scientists have long used such tissues to study the origins of diseases that begin before birth and to develop potential treatments (see table). Fetal cell lines are also used to produce some vaccines. The recordings sparked accusations that Planned Parenthood was illegally profiting from the transactions; federal law prohibits the sale of fetal tissue. Planned Parenthood maintains that it charges recipients only for processing and shipping costs, which is allowed.

Those federal rules don't go far enough for abortion opponents. In states including California and Illinois, lawmakers have introduced or begun to discuss bills that would either revoke public funding for clinics that supply fetal tissue to researchers, exclude such tissue from state laws governing organ donation, or criminalize its use in research.

In Wisconsin, the Planned Parenthood controversy revived a proposal championed for several years by Republican state Representative André Jacque. It targets abortion providers and researchers who are "negotiating and brokering for these body parts," Republican Representative Joel Kleefisch, chair of the Wisconsin Assembly Committee on Criminal Justice and Public Safety, said at a hearing on the bill last week. The measure, which Kleefisch's panel approved on 9 September by a 7 to 4 vote, would make it a felony violation for an individual to "knowingly acquire, provide, receive, or use" cells

NIH's fetal tissue portfolio

HIV research dominated the $76 million in federal funding for studies involving human fetal tissue in 2014 (the most recent available data).



CREDITS: (PHOTO) AMERICAN LIFE LEAGUE/FLICKR; (DATA) NIH

or tissues from a fetus aborted after 1 January 2015. The punishment would be equivalent to that specified for trafficking human organs: a fine of up to $50,000, regardless of whether money changed hands.

Wisconsin's large and vocal biomedical research community has been speaking out against the proposal. The Wisconsin Technology Council, a nonpartisan legislative advisory group that includes members from tech companies, investment firms, and universities, warned that the restrictions would drive out businesses and research talent. And on 26 August more than 700 faculty members at the University of Wisconsin (UW), Madison, signed a letter of opposition to an original draft of the bill, which applied to all fetal tissue regardless of when the abortion was performed.

The letter predicted that the bill would disrupt research in many of the roughly 100 UW laboratories that use cell lines derived from fetal tissue, many of which originated decades ago. That argument prompted the bill's sponsors to rewrite it, so that it applies only to tissue from abortions performed in 2015 or later.

The compromise would still bar many types of research that use new fetal tissue. UW developmental neurobiologist Anita Bhattacharyya, for instance, uses fresh fetal tissue to investigate how abnormal brain development leads to intellectual disability in Down syndrome and fragile X syndrome. The prenatal period is "the time when things are going wrong," she explains. Much of her work now relies on induced pluripotent stem (iPS) cells—adult cells that have been reprogrammed to an embryonic state. But she needs to compare her experimental iPS cells with true fetal cells in order to verify that they accurately mimic brain development.

Bhattacharyya says she, like most colleagues in her field, gets fetal tissue samples from the NeuroBioBank, a network of tissue banks funded by the National Institutes of Health that operates under extensive ethical guidelines. (Planned Parenthood clinics in Wisconsin do not provide donated tissue to researchers.)

Bhattacharyya will learn later this year whether she'll be able to continue her research in Wisconsin. The fetal tissue bill is expected to pass the Republican-led assembly later this fall, and then go to the state senate, also led by Republicans. Wisconsin's Republican governor, presidential candidate Scott Walker, has expressed outrage over the Planned Parenthood videos but has not taken a public position on the measure.

If the bill becomes law, Bhattacharya says researchers in Wisconsin will eventually run out of older tissue, and some, she predicts, will simply leave the state. ■

DATA CHECK

BEHIND THE NUMBERS

GOP legislators choke on ozone standards

By **Jeffrey Mervis**

Later this month, the U.S. Environmental Protection Agency (EPA) is expected to announce tougher federal air-quality standards for ground-level ozone, a key component of smog. Many Republicans oppose the change, calling it another ill-advised government regulation that will hurt the economy.

A group of medical professionals in Congress has added a new twist to that argument. In a 17 June letter to EPA Administrator Gina McCarthy, 22 Republican legislators say that dropping the current standard of 75 parts per billion (ppb) to between 65 ppb and 70 ppb will not improve public health, as the Obama administration and most health officials maintain. And they say that data back them up.

The letter points out that ground-level ozone levels in the United States have fallen by 18% since 2001, even as the prevalence of asthma has risen by 15%. "This lack of correlation," they argue, refutes EPA's argument that "more stringent ozone standards are necessary to protect public health."

It's a novel argument, based on real numbers. But the doctors are overlooking their training in both medicine and statistics.

Janice Nolen, assistant vice president for national policy at the American Lung Association in Washington, D.C., says it's a mistake to link those opposing trends because asthma has many triggers besides ozone. That means prevalence can rise even as smog levels decline.

"Asthma is a really complicated disease," Nolen says. "Along with one's genetic makeup, there are any number of environmental factors that play a role in" susceptibility, such as exposure to pollen, cold air, and dust mites. Just this month, for example, a new study suggested that children exposed to bacteria in dairy farm dust may have suppressed inflammatory reactions, reducing their chance of developing asthma (*Science*, 4 September, p. 1034).

Nolen points out that other data leave no question about the importance of ozone. "A smog alert leads to more hospital admissions," she notes.

EPA's McCarthy says reducing those admissions is one goal of the new standard. The current National Ambient Air Quality Standards for ground-level ozone were set at 75 ppb in 2008 by the Bush administration, which disregarded a recommendation from an external scientific advisory panel to lower it to 60 ppb. In 2011, the Obama administration withdrew a proposal with 60 ppb as a possible target.

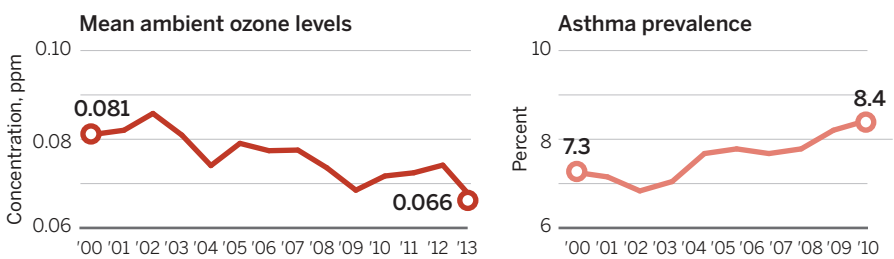
Now, EPA is trying again. McCarthy has said that "stronger ozone standards will provide an added measure of protection for low-income and minority families who are more likely to suffer from asthma." And she doesn't buy the letter's correlational critique. "While I respect their comments and will certainly take them into consideration," she told *Science* this summer, "the agency is going to rely on tried-and-true science."

The Republican legislators declined numerous requests to explain the arguments they made in their letter, which was initiated by Senator Bill Cassidy (R–LA) and Representative Mike Burgess (R–TX). Signers include all 18 members of the House of Representatives's GOP Doctors Caucus.

As medical professionals, the legislators have probably told their patients about the importance of getting a second opinion. Their constituents might want to do likewise on the issue of setting ozone standards. ■

Correlation doesn't mean causation

A rise in asthma prevalence (right) as ozone levels are dropping (left) doesn't invalidate the argument that tougher ozone standards would improve health.



DATA: (LEFT TO RIGHT) OZONE IS ENVIRONMENTAL PROTECTION AGENCY; ASTHMA IS CENTERS FOR DISEASE CONTROL AND PREVENTION/NATIONAL CENTER FOR HEALTH STATISTICS

OPTICS

Skintight invisibility cloak radiates deception

Differing from previous clunky cloaks, new device erases an object's optical signature

By **Adrian Cho**

In J. K. Rowling's blockbuster novel *Harry Potter and the Philosopher's Stone*, the young wizard Harry receives an invisibility cloak: a silky garment that makes him disappear. Now, researchers in the budding field of invisibility cloaking have made an ultrathin cloak of their own, described on page 1310. It works not by magic, but by erasing the signature of an object from reflected light. And it won't hide Harry; it conceals objects just a few micrometers in size. But such a device might be scaled up to hide larger things, some researchers say.

"This cloak really is thin—it is more of what people think of as a cloak," says John Pendry, a theorist at Imperial College London, who in 2006 laid out the theory for the first invisibility cloaks (*Science*, 26 May 2006, p. 1120). Those devices were large compared with the objects they hid—"less like Harry Potter's cloak and more like Harry Potter's shed," Pendry quips. Andrea Alú, an electrical engineer at the University of Texas, Austin, who proposed the new type of cloak, says, "It's a great result."

The original invisibility cloaks smoothly funnel light or longer wavelength electromagnetic waves around an object so that it can't be seen. The first cloak, which worked at only one wavelength, was a ring that guided light around an object inside. A broader band "carpet cloak" consists of a prismlike triangle sitting on a flat surface. It guides light shining from above around a hiding space hollowed out of the bottom of the triangle that resembles a hump in a rug. The light then appears to simply reflect off the underlying surface, rendering anything inside the hiding space invisible.

Such cloaks have limitations. They are made of metamaterial, an assemblage of gizmos smaller than the wavelength of the electromagnetic waves that work en masse like a continuous material with tunable optical properties. To funnel light smoothly, the shell of metamaterial must be much thicker than the wavelength of the radiation.

Instead of funneling light around an object, Xiang Zhang, an applied physicist at the University of California, Berkeley, and his team change light waves bouncing off the object to eliminate traces it would otherwise leave in the waves. As with a carpet cloak, the object sits on a surface (see diagram). Ordinarily, when light's straight, evenly spaced wave fronts strike the object, they are distorted and reflected in all directions—which makes the thing visible. The new cloak counteracts the distortions so the reflected waves remain straight, as if reflected by the underlying surface.

To do that, the researchers apply a "metasurface"—a skintight layer of transparent insulator decorated with rectangles of gold of different sizes—generally less than a micrometer on each side. Each gold patch absorbs light and reradiates it like a tiny antenna. Depending on its dimensions, a patch can shift the undulations of the outgoing wave relative to those of the incoming wave—just as the motion of children on two adjacent swings can be adjusted to be in or out of sync. That "phase shift" comes about because an antenna is either slightly too big or slightly too small to fully ring or "resonate" with light of the incident light. By tweaking the sizes and phase shifts of the antennas, the researchers make all of them radiate in concert to produce the undistorted wave fronts.

To prove the "skin cloak" works, the researchers made one to hide a landscape of random bumps a few micrometers wide and about a micrometer tall. The cloak was 80 nanometers thick, one-ninth the wavelength of the red light used in the experiment.

The skin cloak has its own limitations. The pattern of antennas must be precisely tailored to the bumps beneath it. So you could imagine wearing the new cloak, Zhang says, "but you can't move or it won't work anymore." Also, the cloak works over a limited range of wavelengths—although Zhang says the range is wide enough to encompass most of the visible spectrum.

More important, the features being hidden cannot be too sharp and too big compared with the wavelength of light, says Zeno Gaburro, a physicist at the University of Trento in Italy who helped pioneer metasurfaces. Otherwise, they will cast shadows that a metasurface can't erase. "The face that is dark does not see the light, so there's no way you can correct for [the shadow] with this technique," Gaburro says. Zhang is more sanguine about scaling up the cloak. "I don't see any roadblocks," he says.

Regardless, the paper marks a significant advance, researchers say. "It really proves the power of the metasurface concept," Alú says. Pendry says such surfaces could quickly find practical uses. For example, it might be possible to cover a fighter plane with a metasurface designed for radar, not to make it invisible but to make it look like a freighter. "It might be easier to make one thing look like something else," he says, "and that might be good enough." ■

The skinny on the skin cloak

When light waves strike an object (top), the reflected waves are distorted (middle). The cloak's antennas compensate for the distortions, hiding the object.

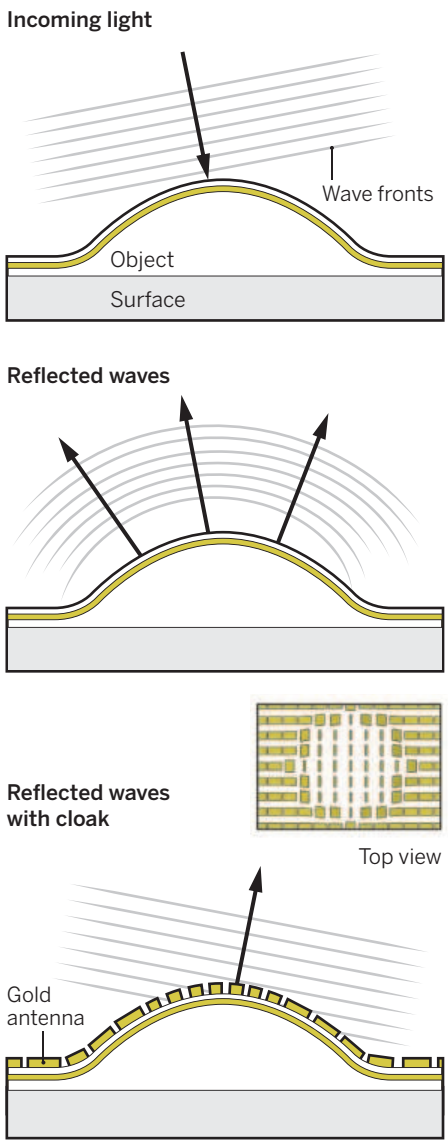


DIAGRAM: V. ALTOUNIAN/*SCIENCE*

HUMAN EVOLUTION

Humanity's long, lonely road

Oldest ancient nuclear DNA suggests humans and Neandertals parted ways early

By **Ann Gibbons**, *in London*

The oldest nuclear DNA ever sequenced from a human ancestor has shaken up our family tree. At a meeting of the European Society for the study of Human Evolution last week, paleogeneticist Matthias Meyer of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, reported that he and his colleagues have analyzed nuclear DNA from mysterious Spanish fossils thought to be 300,000 to 400,000 years old. The result settles a long-standing puzzle about the identity of the fossils. It also implies that the ancestors of modern humans parted ways with their archaic relatives hundreds of thousands of years earlier than was thought.

The finding "resolves one controversy—[the fossils] are in the Neandertal clade," says paleoanthropologist Chris Stringer of the Natural History Museum in London. "But it's not all good news, from my point of view: It pushes back the origin of *Homo sapiens*," presenting him and other experts in human origins with a new puzzle.

The thousands of bones and teeth from 28 individuals found in the mid-1990s at Spain's Sima de los Huesos ("pit of bones") cave had always looked a lot like primitive Neandertals, the archaic humans who inhabited ice-age Europe. But they were so old that researchers had classified them as an earlier species, *H. heidelbergensis*, which lived about 600,000 to 250,000 years ago in Europe, Africa, and Asia and was seen as a possible common ancestor of Neandertals and modern humans.

But in 2013, a first glimpse of DNA from the Sima fossils muddied the picture. Max Planck researchers extracted mitochondrial DNA (mtDNA), which is more abundant but less revealing than nuclear DNA in ancient specimens. They found that it resembled that of the Denisovans, another kind of extinct human most closely related to Neandertals, known from DNA from a finger bone from Denisova Cave in Siberia, Russia (see story, right). That unexpected finding prompted the researchers to try to sequence the Spanish fossils' nuclear DNA.

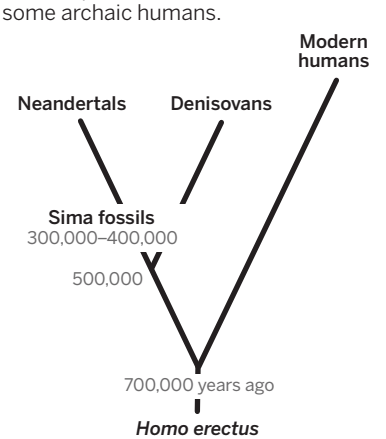
It wasn't easy: The ancient DNA was degraded to short fragments, made up of as few as 25 to 40 single nucleotides. Meyer and his colleagues did not sequence the entire genome of the fossils, but after 2 years of effort they did get 1 million to 2 million base pairs of nuclear DNA—300,000 years older than any reported before from a human ancestor—from a tooth and a leg bone. Sequencing nuclear DNA from such archaic humans "has been the next frontier with ancient DNA," says evolutionary biologist Greger Larson of the University of Oxford in the United Kingdom.

Scanning this DNA, Meyer and his colleagues found that the two Sima fossils share far more genetic markers with Neandertals than with Denisovans or modern humans. "Indeed, the Sima de los Huesos specimens are early Neandertals or related to early Neandertals," Meyer said at the meeting. That suggests that the split of Denisovans and Neandertals should be moved back in time, he added. And because Neandertals and Denisovans were more closely related to each other than to modern humans, the ancestors of modern humans must have split away from the tree even earlier, perhaps 550,000 to 765,000 years ago, Meyer suggested. (All three groups did interbreed at low levels after their evolutionary paths diverged—which may explain the Denisovan-like mtDNA in the first Sima fossil to be sequenced.)

That would mean that the ancestors of humans were already wandering down a solitary path, apart from the other kinds of archaic humans on the planet, 100,000 to 400,000 years earlier than expected. ■

Deeper branches

Putting the Sima fossils on the Neandertal lineage implies an earlier split between modern and some archaic humans.



2015

HUMAN EVOLUTION

Cave was lasting home to Denisovans

Dates and DNA solidify presence of elusive human relative in Siberia

By **Ann Gibbons**, *in London*

In 2010, scientists discovered a new kind of human by sequencing DNA from a girl's pinky finger found in Denisova Cave in Siberia, Russia. Ever since, researchers have wondered when the girl lived, and whether her people, called Denisovans, lingered in the cave or just passed through. But the elusive Denisovans left almost no fossil record—only that bit of bone and a handful of teeth—and they came from a site that was notoriously difficult to date (*Science*, 26 August 2011, p. 1084).

Now, state-of-the-art DNA analysis on the Denisovan molars and new dates on cave material show that Denisovans occupied the cave surprisingly early and came back repeatedly. The data suggest that the girl lived at least 50,000 years ago and that two other Denisovan individuals died in the cave at least 110,000 years ago and perhaps as early as 170,000 years ago, according to two talks here last week at the meeting of the European Society for the study of Human Evolution. Although the new age estimates have wide margins of error, they help solidify our murky view of Denisovans and provide "really convincing evidence of multiple occupa-

CREDITS: (DATA) ADAPTED FROM CHRIS STRINGER, NATURAL HISTORY MUSEUM, LONDON; (PHOTO) MAX PLANCK

Denisovans occupied Denisova Cave repeatedly over more than 100,000 years. Neandertals slipped in as well, and modern humans were the last to live there.

tions of the cave," says paleoanthropologist Fred Spoor of University College London. "You can seriously see it's a valid species."

Most of the cave's key fossils come from a thick band of sandstone called layer 11. When researchers first dated animal bones and artifacts in this layer, the results varied widely, between 30,000 to 50,000 years ago. So Siberian researchers invited geochronologist Tom Higham of the University of Oxford in the United Kingdom to redate the sequence. Higham's team collected and radiocarbon-dated about 20 samples of artifacts and animal bones with cut marks, which presumably were discarded by ancient humans. Sediments holding the finger bone, at the bottom of layer 11, came out right at the limit of radiocarbon dating, and are likely older than 48,000 to 50,000 years, reported postdoc and archaeologist Katerina Douka of Oxford.

Denisovans were discovered when their genome was sequenced from this pinky finger bone from a young girl.

Another dating expert at the meeting was cautious about these results. "How secure is the association of the Denisovans with the [dated] animal remains?" asked geochronologist Daniel Richter of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and Leuphana University in Lüneburg. But Douka stressed that the dates were from cut-marked animal bones and ornaments, and were consistent across three cave chambers.

The dates also fit with genetic evidence presented at the meeting that Denisovans were in the cave early. Researchers sequenced nuclear DNA from three molars from layer 11 and a child's molar from a deeper layer, 22, according to a talk by graduate student Viviane Slon, who works in the lab of paleogeneticist Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology. (A dating method considered experimental for caves, thermoluminescence dating, had suggested that layer 22 is 170,000 years old.)

Slon and her colleagues managed to analyze a significant amount of nuclear DNA from three teeth that turned out to be Denisovan. (A fourth was Neandertal.) By comparing key sites on the tooth DNA with corresponding sites in the high-quality genomes of the Denisovan girl, Neandertals, and modern humans, they revealed that the Denisovan inhabitants in that one cave were not closely related. They had more genetic variation among them than all the Neandertals so far sequenced, although Neandertals are known to be similar genetically.

To find out when the Denisovans were in the cave, the team also sequenced their entire mitochondrial DNA (mtDNA) genomes and placed them on a family tree. Then they counted the number of mtDNA differences between individuals and used the modern human mutation rate to estimate how long it might have taken those mutations to appear. They concluded that the girl with the pinky finger was in the cave roughly 65,000 years after the oldest Denisovan, who was there at least 110,000 years ago and possibly earlier.

Neandertals were in Denisova Cave, too—Pääbo's team has sequenced their DNA from a toe bone and molar found there. And modern humans also were apparently drawn to the large, light-filled cave, to judge from more recent artifacts found there. "What I found fascinating is the interdigitization of the Neandertals and Denisovans—that both groups were in and out of the cave," says paleoanthropologist Leslie Aiello of the Wenner-Gren Foundation in New York City.

The challenge now is to get more fossils to flesh out the still-mysterious Denisovans. To that end, Oxford grad student Samantha Brown reported in a poster that she discovered a human bone fragment by using a new technique, called ZooMS, to scan 2315 bones from the cave for uniquely human proteins. Anything else she finds will be welcome. "This is a real lineage, and we have to work out what the hell it looks like," says paleoanthropologist Bernard Wood of George Washington University in Washington, D.C. ■

PHOTO: MAX PLANCK

ASTRONOMY

India to put observatory in orbit

Suite of imagers will complement other space telescopes

By **Pallava Bagla**, *in New Delhi*

On 28 September, India plans to launch a major space-based observatory that will set a milestone for a developing country—and also mark a first for space science. The $45 million Astrosat mission will put India in an elite group of science powers—the United States, the European Union, and Japan—doing astronomy from orbit. The spacecraft also has a unique design. Instead of focusing on a single region of the spectrum, it carries a suite of five instruments, including two telescopes, which will observe cosmic objects simultaneously in several wavebands: visible, ultraviolet, and soft and hard x-rays.

Astrosat's capabilities are "unique and very powerful," especially for active galactic nuclei and certain x-ray sources, which vary on fast timescales and radiate over many wavelengths, says mission collaborator John Hutchings, an astrophysicist at the Herzberg Institute of Astrophysics in Victoria, Canada. The spectral symphony should allow Astrosat "to unveil a few mysteries of the universe," adds Sujan Sengupta, an astrophysicist at the Indian Institute of Astrophysics in Bengaluru, who is not affiliated with the project.

The launch, from Sriharikota spaceport in Andhra Pradesh, culminates a 2-decade odyssey. A team led by former Indian Space Research Organisation (ISRO) Chairman K. Kasturirangan conceived the mission in 1996; the space agency aimed to put the satellite in orbit in 2007. But funds were slow in coming, and ISRO and its partner institutes encountered hardware-related glitches. The soft x-ray imaging telescope, manufactured at the Tata Institute of Fundamental Research in Mumbai, was especially tricky; hundreds of gold-coated mirrors for reflecting the x-rays onto a charge-coupled device imager had to be positioned to an accuracy of less than the width of a human hair.

Astrosat's multiwavelength capability could pay special dividends with sources that flare up rapidly, such as x-ray binaries. These

pairs, which harbor a dense object such as a black hole and an ordinary star, are undetectable—except when they suddenly flare, or outburst. Observing in wavelengths longer than x-rays can help researchers get a handle on the size and shape of the accretion disks responsible for emissions. In the past, astronomers have marshaled multiple ground- and space-based telescopes to observe outbursts across the spectrum, but doing so entails "a lot of coordination and difficulty," says K. Suryanarayana Sarma, Astrosat's project director at ISRO in Bengaluru. It's hard to free up telescopes on short notice, meaning observations often lag the initial burst.

"To study these sources, it is necessary to detect them as early as possible in the outbursting phase," which lasts only a few months, says Somasundaram Seetha, program director in ISRO's space science program office. The new spacecraft should make that possible by lying in wait, tracking x-ray emissions across the sky. When it catches an outburst, she says, Astrosat's instruments will swing into action, and ISRO will also send out a worldwide alert.

In another first, a U.S. firm will hitch a ride on an Indian launch. Piggybacking on Astrosat's heavy lift rocket are four LEMUR CubeSats, designed for Earth observation and as a technology demonstration by Spire Global, Inc., based in San Francisco, California.

Astrosat's anticipated mission life is 5 years. Observing time will be opened to the global community 1 year after launch, says Kiran Kumar, chairman of ISRO. Kasturirangan is confident that the long delay has not dimmed the observatory's prospects. "There is no doubt," he says, "that Astrosat is still very relevant for the world." ■

Pallava Bagla is a science journalist in New Delhi.

Astrosat is jam-packed with telescopes and imagers.

PHOTOS: (TOP TO BOTTOM) ISRO; IDRISSA SOUMARE/AFREECOM

CLINICAL TRIALS

Ebola vaccines face daunting path to approval

Triumph in a clinical trial is no guarantee a vaccine stock will be available in the next outbreak

By **Jon Cohen** *and* **Martin Enserink**

No vaccine in history has moved forward more quickly than the one against Ebola made by Merck. The first person received the vaccine last fall; in July, 9 months later, an unusual clinical trial conducted in Guinea under the most unforgiving conditions—remote villages, suspicion of researchers, dwindling cases—proved that the shots actually worked. The estimated efficacy was between 74.7% and 100%—such solid protection that the control arm of the study was abandoned and all participants were offered the vaccine (*Science*, 7 August, p. 569). Many say the Merck vaccine may have helped curb the epidemic in Guinea, and it is now also being used in Sierra Leone, in an extension of the trial.

But will this remarkable efficacy translate into a desperately needed stockpile of vaccines ready to quash the next outbreak? The answer will depend on whether developers can persuade regulatory agencies including the U.S. Food and Drug Administration (FDA) that their vaccine is safe and effective enough for widespread use. The Merck vaccine and a different one made by London-based GlaxoSmithKline (GSK), which has not yet proven itself in a clinical trial but looks promising, face multiple hurdles. The unconventional clinical trial of the Merck vaccine may not have generated enough data to satisfy regulators, acknowledges Mark Feinberg, who headed the vaccine project until he left Merck on 7 September. And with the epidemic finally on the wane, collecting more convincing efficacy data on either vaccine has become practically impossible. After 22 gruesome months, more than 28,000 cases, and 11,291 deaths, Liberia is officially Ebola-free, and during the week ending on 6 September, Guinea and Sierra Leone only had one case each.

There are sure to be future outbreaks, making it a "huge priority" to approve an Ebola vaccine so that it can be stockpiled and deployed fast, says Luciana Borio, the acting chief scientist at FDA in Silver Spring, Maryland. "A vaccine that's proven safe and effective for Ebola would be a monumental event," adds Borio, who says the agency will speedily review any application manufactur-

Baldé Thierno Boubacar, a hospital worker in Conakry, in Guinea, takes part in a study of Merck's vaccine.

ers submit. But neither Merck nor GSK has yet filed anything to FDA or key regulatory agencies in other countries. Both companies are gathering more data in human and animal trials, hoping to take advantage of approval routes that do not require massive, controlled trials in the midst of an outbreak.

The Merck vaccine, which contains the gene for an Ebola virus surface protein stitched into a harmless livestock pathogen, vesicular stomatitis virus (VSV), was originally developed in 2003 by the Public Health Agency of Canada. The Canadian government then licensed the vaccine to NewLink Genetics, a small biotech in Ames, Iowa, which relied on funding from the U.S. Department of Defense to develop it. The company had difficulty drumming up interest in the vaccine until the outbreak hit. But last November, when it became clear that hundreds of thousands of doses might be needed in short order, NewLink licensed the vaccine to Merck.

After the VSV-based vaccine quickly passed muster in safety studies, the World Health Organization began the unusual "ring vaccination" trial—a strategy reminiscent of the one used to eradicate smallpox by creating a ring of immunity around infected people. The concept had never been tried in a clinical trial. Standard efficacy studies randomly assign participants to receive either a vaccine or a placebo; in this study, researchers gave the Merck vaccine to clusters of people in Guinea who might have been exposed to a confirmed Ebola case. They then compared Ebola cases in these clusters, which averaged about 80 people each, with those in similarly sized clusters of people who had also potentially been exposed but were randomly assigned to receive the vaccine 3 weeks later. The delayed arm essentially served as a control group. Ten days after receiving one shot, no one in the 48 clusters of immediately vaccinated people had developed Ebola, compared with 16 cases in the 42 delayed clusters.

The study originally had planned to compare 95 immediate clusters with 95 delayed ones, and the early termination of the control arm of the study decreased the statistical "power" of the results. But if the data are not robust enough to obtain FDA's standard efficacy approval, Merck could pursue either of two alternative approval processes in place for drugs and vaccines designed for "serious or life-threatening conditions" that for practical or ethical reasons cannot meet the traditional efficacy requirements. So could GSK, whose vaccine uses a chimpanzee adenovirus with the Ebola surface protein gene.

Both vaccines already appear to satisfy the first of these, the animal rule: They have proven to be safe in humans and have protected vaccinated monkeys that were injected with the virus to "challenge" their immunity. But FDA will not consider the animal rule if a vaccine possibly can meet what's known as the accelerated approval requirements, which fall midway between demonstrating formal efficacy and the animal rule. Accelerated approval has been used for drugs but not vaccines, so when FDA suggested it in May, "a light bulb went off," says Ripley Ballou, head of the program developing GSK's Ebola vaccine. "I don't think anyone had thought of it before."

In essence, accelerated approval demands that researchers determine which immune responses protect vaccinated monkeys and then show that the product triggers similar responses in humans. To figure out which immune responses in vaccinated monkeys correlate with protection, researchers compare animals that get sick after a viral challenge with those that stay healthy. But both the GSK and Merck vaccines had 100% success in monkey tests, which makes it difficult to tease out the immune mechanisms behind their efficacy. To find the correlates of immunity, both companies are now giving lower doses of the vaccines to monkeys before challenging them, which theoretically should reveal the immunologic breaking points between success and failure.

Both companies also want more data about the immune responses in vaccinated humans. A massive trial planned in Liberia of both vaccines was suspended because the epidemic there ended; researchers plan to take blood from the people enrolled in the study—500 for each of the two vaccines—for a full year. Similar studies of the Merck vaccine alone are taking place in Guinea and Sierra Leone. And GSK in June launched new studies in five West African countries near this epidemic—three of which had spillover Ebola cases—to assess safety and immune responses.

At this point, FDA says it remains open to all three approval options. "Each and every one of these pathways has uncertainties," Borio says. But if the companies receive licensure via the animal rule or accelerated approval, FDA will require postmarketing studies in future outbreaks. Just what those would look like "is not entirely clear," Feinberg says. It would be unethical to withhold a licensed product to create a control group, and if more than one vaccine receives approval, trials become more complex still.

If another outbreak erupts in the interim, countries may well request to use one or both of these vaccines experimentally. But a far better option is to have an approved vaccine in hand at the start of an outbreak that can be used in conjunction with traditional containment efforts such as isolation of cases, contact tracing, and safe burial, says Clifford Lane of the U.S. National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, who is running the Liberian vaccine study. "If things are able to be mobilized quickly enough," Lane adds, "we probably won't find ourselves in the situation that we're just leaving." ■

The winding road to an Ebola vaccine

These phase II and phase III trials may yield additional information needed for regulatory approval.

VACCINES/LOCATION	TARGET ENROLLMENT	START DATE	DESIGN	STATUS
Merck, GSK/Liberia	27,000 in general population	February 2015	Randomized controlled with placebo arm	Stopped at 1500. Blood collection continues
Merck/Sierra Leone	8700+ frontline workers	April 2015	Immediate versus deferred	Immediate arm vaccinated, deferred in fall
Merck/Guinea	190 clusters of potential contacts	April 2015	Ring vaccination, immediate versus deferred	Control arm halted after analysis of first 90 clusters, all offered vaccine
Merck/Guinea	1200 frontline workers	March 2015	Safety and immune responses	May add 2000 more
GSK/Mali, Senegal, Ghana, Cameroon, Nigeria	3000 adults	June 2015	Safety and immune responses	Plan to add 600 children in October

DATA: NIH, CDC, MSF, WHO, GSK, MERCK

A TRIAL FOR THE AGES

Nir Barzilai wants to launch the first rigorous test of a drug that could put the brakes on aging

By **Stephen S. Hall**

On a blazingly hot morning this past June, a half-dozen scientists convened in a hotel conference room in suburban Maryland for the dress rehearsal of what they saw as a landmark event in the history of aging research. In a few hours, the group would meet with officials at the U.S. Food and Drug Administration (FDA), a few kilometers away, to pitch an unprecedented clinical trial—nothing less than the first test of a drug to specifically target the process of human aging.

"We think this is a groundbreaking, perhaps paradigm-shifting trial," said Steven Austad, chairman of biology at the University of Alabama, Birmingham, and scientific director of the American Federation for Aging Research (AFAR). After Austad's brief introductory remarks, a scientist named Nir Barzilai tuned up his PowerPoint and launched into a practice run of the main presentation.

Barzilai is a former Israeli army medical officer and head of a well-known study of centenarians based at the Albert Einstein College of Medicine in the Bronx, New York. To anyone who has seen the ebullient scientist in his natural laboratory habitat, often in a short-sleeved shirt and always cracking jokes, he looked uncharacteristically kempt in a blue blazer and dress khakis. But his practice run kept hitting a historical speed bump. He had barely begun to explain the rationale for the trial when he mentioned, in passing, "lots of unproven, untested treatments under the category of anti-aging." His colleagues pounced.

"Nir," interrupted S. Jay Olshansky, a biodemographer of aging from the University of Illinois, Chicago. The phrase "anti-aging … has an association that is negative."

"I wouldn't dignify them by calling them 'treatments,'" added Michael Pollak, director of cancer prevention at McGill University in Montreal, Canada. "They're products."

Barzilai, a 59-year-old with a boyish mop of gray hair, wore a contrite grin. "We know the FDA is concerned about this," he conceded, and deleted the offensive phrase.

> ***"What we're talking about here is a fundamental sea change in how we look at aging and disease."***
>
> **Jay Olshansky**, University of Illinois, Chicago

Then he proceeded to lay out the details of an ambitious clinical trial. The group—academics all—wanted to conduct a double-blind study of roughly 3000 elderly people; half would get a placebo and half would get an old (indeed, ancient) drug for type 2 diabetes called metformin, which has been shown to modify aging in some animal studies. Because there is still no accepted biomarker for aging, the drug's success would be judged by an unusual standard—whether it could delay the development of several diseases whose incidence increases dramatically with age: cardiovascular disease, cancer, and cognitive decline, along with mortality. When it comes to these diseases, Barzilai is fond of saying, "aging is a bigger risk factor than all of the other factors combined."

But the phrase "anti-aging" kept creeping into the rehearsal, and critics kept jumping in. "Okay," Barzilai said with a laugh when it came up again. "Third time, the death penalty."

The group's paranoia about the term "anti-aging" captured both the audacity of the proposed trial and the cultural challenge of venturing into medical territory historically associated with charlatans and quacks. The metformin initiative, which Barzilai is generally credited with spearheading, is unusual by almost any standard of drug development. The people pushing for the trial are all academics, none from industry (although Barzilai is cofounder of a biotech company, CohBar Inc., that is working to develop drugs targeting age-related diseases). The trial would be sponsored by the nonprofit AFAR, not a pharmaceutical company. No one stood to make money if the drug worked, the scientists all claimed; indeed, metformin is not only generic, costing just a few cents a dose, but belongs to a class of drugs that has been part of the human apothecary for 500 years. Patient safety was unlikely to be an issue; millions of diabetics have taken metformin since the 1960s, and its generally mild side effects are well-known.

Finally, the metformin group insisted they didn't need a cent of federal money to proceed (although they do intend to ask for

PHOTO: RICK WENNER

Barzilai hopes to persuade FDA to bless the proposed anti-aging trial, which is unconventional in its goals and design.

some). Nor did they need formal approval from FDA to proceed. But they very much wanted the agency's blessing. By recognizing the merit of such a trial, Barzilai believes, FDA would make aging itself a legitimate target for drug development.

By the time the scientists were done, the rehearsal—which was being filmed for a television documentary—had the feel of a pep rally. They spoke with unguarded optimism. "What we're talking about here," Olshansky said, "is a fundamental sea change in how we look at aging and disease." To Austad, it is "the key, potentially, to saving the health care system."

As the group piled into a van for the drive to FDA headquarters, there was more talk about setting precedents and opening doors. So it was a little disconcerting when Austad led the delegation up to the main entrance of FDA—and couldn't get the door open.

Barzilai, who is quick to see the humor (black or otherwise) in any situation, leaned over to a colleague and muttered, "I hope it's not symbolic."

AMIABLY AGGRESSIVE, Barzilai credits his military service with shaping his scientific temperament and administrative mettle. "The most helpful years in my life were in the Israeli army," he says. Born in Haifa in 1955, he joined the military as a medical instructor and in 1976 served as a medical officer in the special forces raid on Entebbe, Uganda, that freed 102 hostages from a hijacked Air France airliner. ("I did nothing," he claims. "I happened to be there.") By his early 20s, he had been decorated for his military service and had become chief medic of the Israeli army. "I had an office, a secretary, a car, and I would do inspections with a helicopter," he recalls. "It's a whole lifetime, okay? With a lot of lessons. But the major thing is that you realize that you can do a lot! If it doesn't frighten you, you can do a lot."

After obtaining his M.D. in internal medicine from the Technion-Israel Institute of Technology in 1985, Barzilai focused on endocrinology and metabolism during several research stints in the United States. He had every intention of returning to Israel, but a fellowship at Yale University in 1987–88 upended his plans. At Yale, he studied the mechanism of several oral drugs that lowered blood sugar in diabetics. One was metformin. During his time in New Haven, Barzilai ended up meeting his wife-to-be on a blind date. Metformin, he laughs, "is the reason I'm in the United States!"

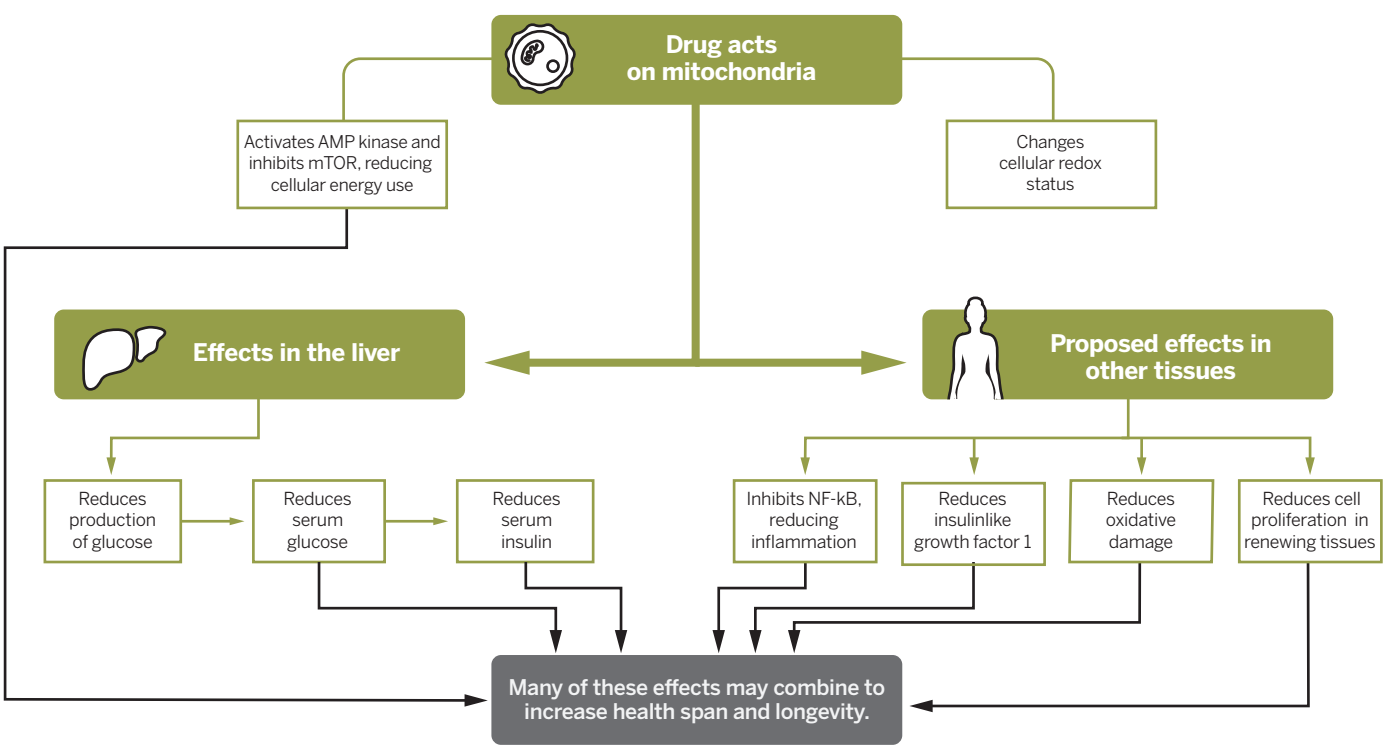
In 1993, he landed a position at Einstein and discovered, in nearby neighborhoods, a cohort of unusually long-lived Ashkenazi Jews. Barzilai and his Einstein colleagues tracked down and studied more than 600 of the "oldest old," ranging in age from 95 to 112 years. Since then, the researchers have collected data on the genetics, epigenetics, blood chemistry, and lifestyles of both centenarians and their children.

The Einstein team has identified, for example, variants in two genes associated with increased levels of high density lipoproteins (which have a protective effect on cardiovascular health in this population). They have found that centenarians more often carry a genetic variant that down-regulates insulinlike growth factor 1 (IGF-1), part of a hormonal pathway that not only regulates growth, but also has been implicated in aging. They have also found that the families of centenarians produce unusually high levels of a peptide created in mitochondria, the cell's power plants, and then released into the body; called "humanin," it is one of several related mitochondrial proteins that may protect against aging. (Barzilai and collaborator Pinchas Cohen of the University of California, Los Angeles, have puckishly dubbed them "Schleps," for "small humaninlike peptides," or SHLPs.)

The study of the Bronx centenarians built on a surge of basic aging research in model organisms like yeast, fruit flies, and

Double dividends

Metformin acts on the mitochondria, the cellular power plants. The result is two sets of effects, one in the liver that explains the drug's benefits in diabetes and the other, less well understood, that could slow aging.



CREDITS: (DIAGRAM) G. GRULLÓN/*SCIENCE*; (DATA) MICHAEL POLLAK, MCGILL UNIVERSITY AND JEWISH GENERAL HOSPITAL, MONTREAL

nematodes. By manipulating individual genes and measuring effects on life span, researchers could test the role of specific molecular pathways in aging. In perhaps the most dramatic mammalian example, Andrzej Bartke, a biologist at the Southern Illinois University School of Medicine in Springfield, showed that mice with mutated growth pathways, which disabled both growth hormone and IGF-1, were much smaller—but lived much longer.

Within the last decade or so, researchers have settled upon what Felipe Sierra, director of the division of aging biology at the National Institute on Aging (NIA), calls "the major pillars of aging." These pathways and mechanisms, roughly half a dozen in all, affect metabolism, growth, response to stress, stem cell vigor, inflammation, and proteostasis—the cell's quality control system for proteins. And their identification has opened the door to a previously outlandish notion. It "allows us to think that, okay, if we understand how this happens, we can maybe manipulate it," Sierra says.

MORE THAN A DECADE AGO, Barzilai and others began lobbying FDA to consider drugs that might do just that. But those discussions bogged down, he says, after the sides couldn't agree on the kinds of biomarkers associated with aging that could be quantified and tracked during a clinical trial.

Barzilai now believes the answer is to design a drug trial that, rather than targeting aging per se, tries instead to delay the onset of "comorbidities": the chronic diseases whose incidence rises sharply as people get older. "Basically, I think the FDA will be more willing to accept something called 'comorbidities' than it is to accept something called 'aging,'" Barzilai says. "Even in our mind, in my mind, aging is not a disease," he adds. "It's, you know, humanity! You're born, you die, you age in between … I'm kind of saying, 'I don't care what they want to call it, if I can delay it.'"

The comorbidity strategy is key to a concept known as the "longevity dividend," first proposed by a group of public policy and health care experts in 2006. The idea is that slowing down the process of aging, even modestly, would have enormous benefits for quality of life and the economics of health care. "We're not arguing—and we've never argued—that we're trying to achieve life extension," says Olshansky, who has pushed the concept while criticizing some of the more outlandish claims in the aging field, such as British gerontologist Aubrey de Grey's prediction that human life spans of 1000 years are possible. "We'll probably live a little longer if we succeed, but that's not the goal," Olshansky says. "The goal is the extension of the period of healthy life."

Even a modest delay in aging could increase average life expectancy by 2.2 years, compress the period of morbidity at the end of life, and save perhaps $7.1 trillion in health care costs over a period of 50 years, Olshansky and colleagues estimated in a 2013 paper in the journal *Health Affairs*. To achieve those benefits, "we've got to act quickly," he argues. "The numbers of people that are frail and disabled [are] rising fairly rapidly, and we're seeing an increase in unhealthy life span."

But the FDA drug approval process abides by the "one disease, one drug" model. Would the agency be open to a trial that had multiple illnesses as an endpoint? As an initial step, earlier this year Sierra organized seminars at FDA in which NIA researchers described recent findings in the biology of aging. In May, Robert Temple, deputy director of FDA's Center for Drug Evaluation and Research, spoke at an NIA retreat.

Encouraged by the tenor of these discussions, Barzilai and a core group of collaborators—Einstein's Jill Crandall; Austad; Olshansky; Stephen Kritchevsky at Wake Forest School of Medicine (where the multicenter trial would probably be based); and James Kirkland, a diabetes researcher at the Mayo Clinic, among others—pushed ahead with plans for the trial.

The next question was: What would be the best drug to test?

THERE WAS NO SHORTAGE of possibilities. Buoyed by the advances in basic research, NIA in 2003 inaugurated a program of animal experiments to test compounds that might alter or slow down the aging process. NIA-supported researchers have tested 16 compounds in mice. Five have shown a positive effect, Sierra says: aspirin, acarbose (a widely prescribed diabetes drug), 17-alpha-estradiol (the nonfeminizing form of estrogen), nordihydroguaiaretic acid (an herbal compound derived from the creosote plant), and the immunosuppressive drug rapamycin (used in organ transplant recipients). (Among the compounds that had no impact are fish oil, green tea extract, curcumin, and the much-ballyhooed red wine ingredient resveratrol.) Rapamycin was the most impressive. "It has advanced to the point in which we not only know it extends life span," Sierra says, "but more importantly, it extends health span."

Barzilai probed aging factors from by studying centenarians and their children, including Jerome Wiesenberg, 83.

Metformin, the drug the group ultimately decided to take to FDA, was not among the compounds that starred in the animal trials. But it has both a promising history and a long, reassuring track record.

"It all starts in the Middle Ages," says McGill's Pollak. "There were herbalists in Europe—and, independently, herbalists in China—who found plant extracts that were useful when people came in complaining of urinating too much." The extracts derived from a perennial herb (*Galega officinalis*) known variously as goat's rue, French lilac, Spanish sainfoin and false indigo. "It worked for some people," Pollak says. "In retrospect, [we know] the people for whom it was working were diabetic."

It wasn't until the late 1800s that chemists isolated the active ingredient in French

PHOTO: RICK WENNER

lilac—a compound known as guanidine. But guanidine itself proved too toxic to humans, so chemists began to synthesize less toxic analogs known as biguanides, including metformin. In the 1950s, a French physician and pharmacologist named Jean Sterne began to test biguanides in patients with type 2 diabetes at a hospital in Paris. "The best one in terms of efficacy was metformin," Pollak says.

Sterne coined the name glucophage ("glucose eater") when he published his results in 1957, the same year the drug was approved for use in France. Approved in the United Kingdom in 1958 and in Canada in 1972, metformin went on to become the biggest selling diabetes drug in the world. However, U.S. regulators didn't approve it until 1994. (FDA requested additional studies, Barzilai drily notes, "to see if metformin works in the same way as in the United Kingdom, because we are so different here.") By now, companies churn out an estimated 37,000 metric tons of the compound annually, most of it in India.

Metformin has its roots in centuries-old herbal remedies including goat's rue, which helped people with symptoms now recognized as diabetes.

Hints that metformin might also prevent diseases associated with aging began to emerge over the past several decades. In a 1998 report by the United Kingdom Prospective Diabetes Study Group, metformin use not only reduced the risk of all diabetes-related complications (including death) by 32%, but also significantly lowered the risk of cardiovascular disease, including heart attack and stroke. A randomized, placebo-controlled trial called the Diabetes Prevention Program showed similar effects, cutting the onset of type 2 diabetes by 31% in a middle-aged population at high risk of developing the disease.

Epidemiological studies have also suggested that metformin reduces cancer risk and mortality and preserves cognitive function. And in a big-data study that, although observational, got the attention of many aging researchers, British researchers reported late last year that in a retrospective analysis of 78,000 adult type 2 diabetics in their 60s, those who took metformin on average lived longer than healthy age-matched controls.

None of these studies proves that metformin will delay the onset of age-associated diseases, and scientists haven't identified an exact mechanism by which the drug might work. But it appears to act on some of the same molecular pathways identified by basic aging research. Besides its effects on blood glucose, metformin affects multiple pathways involved in growth, inflammation, and metabolism (see graphic, p. 1276).

Pollak has demonstrated what he and others see as the key effect, which may trigger the other benefits of the drug: It inhibits oxygen consumption in mitochondria, in effect turning down the cell's metabolic thermostat. "When a furnace is burning," he says, "it's heating up and it's cracking and it begins to degrade. When you keep your house at a lower temperature, your furnace is going to last longer."

As it turns out, Barzilai is very familiar with metformin—not only as a doctor who has prescribed it and as a researcher who has studied it, but as a patient who has taken it for 5 years. (He says he is considered prediabetic.) He can testify to its safety and tell you exactly how to avoid its most common side effect, gastrointestinal upset. "There's nothing we don't know about metformin," Barzilai says—especially its record for safety, which he calls "critical" to the proposed trial.

His colleagues agreed, sometimes reluctantly. "Rapamycin would have been my first choice, because the animal results have been so spectacular," Austad says. "But Nir said, 'We can't afford in this first trial to kill anybody.' And I thought, 'Strategically, he's right.'"

Barzilai concedes that he and the AFAR-sponsored group are as interested in setting a precedent as in scoring an impressive initial success. Satisfying FDA concerns about a trial that breaks tradition and measures multiple disease endpoints in an aging population, they say, will open the door for pharma to enter the field.

"Metformin is for us a tool—a very exciting tool," Barzilai said prior to the FDA meeting. "It'll work, I think. But I don't want to waste the hour talking about metformin. You know, we chose metformin in order not to talk about it anymore."

When he and the rest of the AFAR delegation finally made it into FDA's meeting room, Barzilai scanned the large contingent seated around the table. "Too many young people here!" he joked. "We should leave now!" But the turnout was encouraging—14 FDA staff members, including Temple and several division chiefs. The meeting ran nearly 30 minutes past the scheduled hour, and by the time Barzilai and the others emerged, they wore surprised smiles. Austad flashed two thumbs up. "I don't think it could have gone much better," he said.

Barzilai, whose enthusiasm occasionally exceeds his command of English, sent out an email the next day to everyone who had helped prepare for the FDA meeting, thanking them and describing the meeting as "hysterical." Historical, Barzilai later explained, because "I think that in their heart, they buy it. Or many of them, or the important people, are buying what we are saying." Olshansky left the meeting convinced that FDA had given a green light, contingent on several adjustments to the protocol, which the group is now making.

Other participants, like Sierra, struck a more cautious note. When asked whether FDA representatives expressed skepticism about the proposed trial, he said, "Conceptually? No. But in the details, yes."

SANDY WALSH, an FDA spokeswoman, says the agency does not comment on drugs under development or under investigation. But in a followup communication to the AFAR group, Barzilai says, FDA indicated that although it is not yet convinced that the proposed trial design can establish that metformin has an anti-aging effect, the agency recognizes the potential value in a drug that could improve quality of life and survival—whether the indication sought is aging or multiple morbidities—and is not opposed to the idea of a trial.

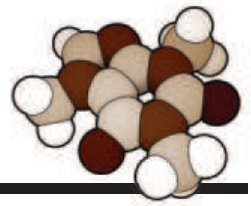
Now, trial advocates need someone willing to foot the cost—"$50 million, plus or minus $20 million," according to Barzilai—of tracking some 3000 people between the ages of 65 and 79 for a minimum of 5 years. Olshansky says the metformin group has already targeted "high net-worth individuals" to bankroll the trial. Federal funding would be welcome, Barzilai says, but private money would probably allow the trial to start sooner. "For me," he says, "the best thing that can happen is that people are writing about it, the television will show it, somebody will call me one day and say: 'You know, I'm rich like I don't know what, and I don't mind helping. Is $50 million enough?' And then we'll get going." ■

Stephen S. Hall is a science writer in Brooklyn, New York.

ILLUSTRATION: BIODIVERSITY HERITAGE LIBRARY

On the renaissance of coal *p. 1286*

How caffeine disturbs sleep *p. 1289* ▸



PERSPECTIVES

ENVIRONMENTAL SCIENCE

Reform forest fire management

Agency incentives undermine policy effectiveness

By **M. P. North,[1,2]* S. L. Stephens,[3] B. M. Collins,[1,3] J. K. Agee,[4] G. Aplet,[5] J. F. Franklin,[4], P. Z. Fulé[6]**

Globally, wildfire size, severity, and frequency have been increasing, as have related fatalities and taxpayer-funded firefighting costs (*1*). In most accessible forests, wildfire response prioritizes suppression because fires are easier and cheaper to contain when small (*2*). In the United States, for example, 98% of wildfires are suppressed before reaching 120 ha in size (*3*). But the 2% of wildfires that escape containment often burn under extreme weather conditions in fuel-loaded forests and account for 97% of fire-fighting costs and total area burned (*3*). Changing climate and decades of fuel accumulation make efforts to suppress every fire dangerous, expensive, and ill advised (*4*). These trends are attracting congressional scrutiny for a new approach to wildfire management (*5*). The recent release of the National Cohesive Wildland Fire Management Strategy (NCWFMS) (*6*) and the U.S. Forest Service's (USFS's) current effort to revise national forest (NF) plans provide openings to incentivize change. Although we largely focus on the USFS, which incurs 70% of national firefighting costs (*7*), similar wildfire policies and needed management reforms are relevant throughout the United States and fire-prone areas worldwide.

POLICY

Accumulated fuels in dry forests need to be reduced so that when fire occurs, rather than "crowning out" and killing most trees, it is more likely to burn along the surface at low-moderate intensity, consuming many small trees and restoring forest resilience to future drought and fire. Mechanical thinning can reduce tree density and some fuels but is often limited by legal (wilderness and park areas), operational (steep or remote ground), and cost constraints (*8*). Fire can also be used to reduce fuels either intentionally (prescribed burning) or opportunistically (letting a natural ignition burn as "managed wildfire") under moderate weather conditions. Although these burns are much less precise than mechanical thinning, in remote locations, fire is usually more efficient, cost-effective, and ecologically beneficial than mechanical treatments (*9*).

Many severe wildfires are due to past fire suppression. Firefighters during the Rim Fire near Yosemite National Park, California, 25 August 2013.

ENTRENCHED DISINCENTIVES. Management reform in the United States has failed, not because of policy, but owing to lack of coordinated pressure sufficient to overcome entrenched agency disincentives to working with fire. Responding to established research, official agency policy now supports a more flexible response to fire than ever before (*6*). Actual wildfire response, however, has changed little because of substantial management impediments. Suppression generally begets larger, more intense wildfires, which in turn intensifies agencies' suppression response (*10*). The alternative, working with fire, is rarely used because of liability and casualty risks and little tolerance for management errors.

PHOTO: © JAE C. HONG/AP/CORBIS

Published by AAAS

For example, during the most recent decade when data were collected (ending in 2008), only 0.4% of ignitions were allowed to burn as managed wildfires (*7*). For individual NFs, there is little economic incentive to change because fire suppression is steadfastly financed through dedicated congressional appropriations, which are augmented with emergency funding, whereas fuels reduction and prescribed burning costs come out of a limited budget allotted to each NF and is often borrowed to cover wildfire suppression costs. With these deterrents, "battling" fire and "only you can prevent wildfire" campaigns have more traction than recognizing that many severe fires result from accrued management decisions. This skewing of agency motivation also distorts economic, insurance, and local regulatory incentives that influence development in fire-prone regions (*11*).

Although agencies are slow to reform internally, they may more rapidly respond to local stakeholder pressure. The core problem has been the lack of a public constituency that advocates for reform of fire-use practices (*11*). The benefits of greater fire use have been a difficult sell because of public objections to smoke and a negative perception of forest fires. This has begun to change as communities increasingly threatened by large fires are urging land-management agencies to accelerate fuel reduction efforts, including the use of managed fire (e.g., yosemitestanislaussolutions.com and 4FRI.org). Timber companies would also benefit from more fire-resilient landscapes in which their private lands are embedded. There is growing awareness that large, severe fires are inevitable in many dry forests, especially in a warming climate. Smoke, safety threats, fire intensity, and human health risks can be better managed for public benefit with proactive fire use under favorable weather and wind dispersal conditions (*12*).

EFFECTING CHANGE. Public support for expanded fire use could thus be directed toward revision of each NF plan, which provides standards and guidelines for daily management decisions. Plans can divide the landscape into zones for different fire management strategies, an approach used by Parks Canada. U.S. forest plans could zone areas close to homes (wildland-urban interface) as an area where most fuels reduction relies on mechanical thinning and fires are suppressed. Beyond this could be an intermediate area where prescribed fire and mechanical treatment are used to optimize fuels reduction. More remote forests could be intentionally burned with prescribed fire, or lightning ignitions allowed to burn as managed wildfires under moderate weather conditions.

Three of the first eight NFs to develop new plans have proposed that more than half of their area in the southern Sierra Nevada be zoned for prescribed and managed fire use. Over the next decade, most of the 155 NFs will begin writing new plans and holding public forums. Engaged local stakeholders will need to look beyond short-term impacts of fire use (e.g., smoke, limited access, and risk of escape) to support managers working with fire and challenge suppression in remote forest zones.

Public support of NCWFMS may help overcome reform disincentives by stressing national interagency collaboration. In response to decades of problem wildfires, the U.S. Congress passed the FLAME Act in 2009 requesting development of NCWFMS, a coordinated strategy to support landscape restoration and fire-adapted communities. Coordination is essential as large, intense wildfires often cross ownership boundaries. For example, in California's 2013 Rim Fire, large patches of old-growth trees in Yosemite National Park were killed when fuel-loaded forests on nearby NF land generated extreme fire behavior that crossed into the park (*13*). NCWFMS can exert peer pressure between agencies and provide support for tough decisions. To accomplish these changes, some policy and resource-deployment decisions supporting fire use could be made at the national level. In the United States, federal land agencies each fund their own fire crews but the National Interagency Fire Center (NIFC) coordinates resource deployment between agencies and nationally across geographic areas. Dedicated crews could be hired and trained for managed fire use, and NIFC could be charged with deploying them for beneficial burning (*14*). Some local and regional agencies have briefly created such crews, but they were often pulled into fire suppression when wildfire activity increased. By giving NIFC deployment authority, it could ensure that these crews are only used for working with fire and are available to burn when weather conditions are favorable. Optimal weather and smoke dispersal conditions occur even in heavily populated and regulated areas such as California, but many burn windows are missed because crews are at or being held for wildfire deployment (*9*). Air-quality regulations limit prescribed fires, although they have much lower emissions than the inevitable wildfire. The Environmental Protection Agency could consider treating prescribed fire smoke like wildfire, as an unregulated "exceptional event."

National government also has an incentive to reduce wildfire expenses and forest agencies' emergency fire borrowing. In many years, suppression costs consume 50% of agency annual budgets, which, after operating expenses, leaves little money for proactive fuels treatment or forest restoration (*11*). Costs and injuries, however, are much lower on managed fires than on escaped wildfires (*7*, *15*). The estimated cost savings for using managed fire compared with wildfire suppression over the same area (*15*) could be reported to Congress to highlight the economy of using proactive restoration rather than reactive triage.

Increased fire use will necessitate management changes (*16*). Mechanical fuels reduction could also be used not only for fire containment but also to establish safe-zone anchors to facilitate greater fire reintroduction (*8*). Large prescribed burns commonly used in Western Australia are possible because a network of these anchors allows 6 to 8% of the forest to be burned annually (*16*). Australian foresters make substantial efforts to educate the public about the inevitability of fire and its ecological benefits and to build support for fire use and smoke tolerance.

We will not eliminate wildfire, but public support for proactive use of managed fires can help restore millions of hectares of forest ecosystems. ■

[1]*USFS Pacific Southwest Research Station, Davis, CA 95618, USA.* [2]*University of California, Davis, CA 95616, USA.* [3]*University of California, Berkeley, CA 94720, USA.* [4]*University of Washington, Seattle, WA 98195, USA.* [5]*The Wilderness Society, Denver, CO 80202, USA.* [6]*Northern Arizona University, Flagstaff, AZ 86011, USA.* **Corresponding author. E-mail: mpnorth@ucdavis.edu*

REFERENCES

1. S. L. Stephens *et al.*, *Front. Ecol. Environ.* **12**, 115 (2014).
2. M. A. Finney, *For. Ecol. Manage.* **211**, 97 (2005).
3. D. E. Calkin *et al.*, *J. For.* **103**, 179 (2005).
4. M. A. Moritz *et al.*, *Nature* **515**, 58 (2014).
5. Staff of Sen. M. Cantwell, "The Wildfire Management Act of 2015: A white paper" (Office of Sen. M. Cantwell, Washington, DC, 2015); http://1.usa.gov/1hDuJ2D.
6. Forests and Rangelands, National Cohesive Wildland Fire Management Strategy, (2014); http://1.usa.gov/1EDCyQL.
7. National Interagency Fire Center, Statistics (NIFC, Boise, ID, 2015); http://1.usa.gov/1NZBOZ4.
8. M. North *et al.*, *J. For.* **113**, 40 (2015).
9. L. N. Quinn-Davidson, J. M. Varner, *Int. J. Wildland Fire* **21**, 210 (2012).
10. D. E. Calkin *et al.*, *For. Ecosyst.* **2**, 9 (2015).
11. K. Bradshaw, D. Lueck, *Wildfire Policy: Law and Economics Perspectives* (Routledge, London, 2012).
12. D. Schweizer, R. Cisneros, *J. Environ. Manage.* **144**, 265 (2014).
13. J. M. Lydersen, M. P. North, B. M. Collins, *For. Ecol. Manage.* **328**, 326 (2014).
14. M. P. Dombeck, J. E. Williams, C. A. Wood, *Conserv. Biol.* **18**, 883 (2004).
15. G. Snider, P. J. Daugherty, D. Wood, *J. For.* **104**, 431 (2006).
16. R. J. Sneeuwjagt *et al.*, *Fire Ecol.* **9**, 14 (2013).

ACKNOWLEDGMENTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the USFS.

10.1126/science.aab2356

GENETICS

Strength in small numbers

A small-scale genome study of an indigenous population elucidates the genetics that influence height and weight

By **Sarah Tishkoff**

Complex traits such as height and weight are influenced by multiple genes and by environmental factors. Because of this complexity, genetic studies of these traits often involve hundreds of thousands of individuals. These studies typically focus on urban populations, for which large sample sizes are readily available. For example, the GIANT consortium recently reported 697 genetic variants that explain 20% of the heritability of adult height (*1*), and 97 loci that account for just 2.7% of phenotypic variation in body mass index (BMI) (*2*). These discoveries required a staggering 250,000 to 300,000 individuals in total. And yet, on page 1343 of this issue, Fumagalli *et al.* report the identification of genetic variants influencing height and BMI based on data from just a few thousand Inuit individuals (*3*). These results exemplify how modern humans have adapted to diverse climates and diets.

The authors increase the statistical power for identifying genetic associations by first searching for targets of natural selection and then looking for association between those loci and potentially adaptive traits. Thus, they first identified genomic targets of natural selection in 191 individuals of Inuit ancestry residing in Greenland and then looked for an association of those regions with anthropometric and metabolic traits in a larger, but still modest, set of ~4500 Inuit individuals.

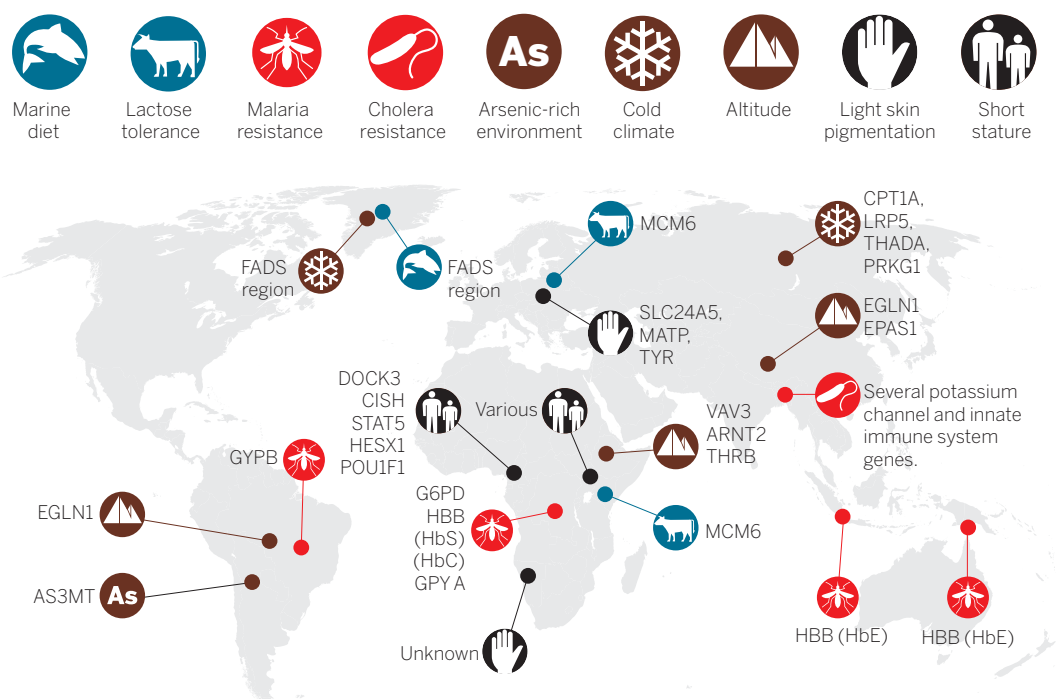
The Inuit population has inhabited the Arctic region for thousands of years, adapting to a cold climate and a diet rich in marine-derived fat. Using a single-nucleotide polymorphism (SNP) genotyping array designed to target metabolism-related genetic variants [the Metabochip (*4*)] and exome sequencing of a subset of 18 individuals, Fumagalli *et al.* identified several variable genome regions that differ in frequency in the Inuit population compared to Europeans and Chinese, possibly due to long-standing selective pressures to adapt to the Arctic environment. Genes found in these regions include those that code for fatty acid desaturases (FADS, important modulators of fatty acid composition), in addition to genes that play a role in fat distribution and in muscle and heart development. Derived variants that were targets of selection near the FADS loci were associated with smaller body size and shorter stature in the Inuit cohort. Two of these variants were modestly associated with a decrease in height in a cohort of 170,000 Europeans. These variants had not been significantly associated with height or body size in the larger European cohort, possibly because of their low frequency in that population.

Fumagalli *et al.*'s study shows that inclusion of ethnically diverse populations that have adapted to extreme environments for thousands of years may be particularly informative for genetic studies of variable traits. The analysis of genetic variants in individuals with phenotypes at the extremes of the trait distribution has been successfully used to identify functionally relevant genetic variants within populations. For example, analysis of European-descent individuals with extremely low concentrations of low-density lipoprotein (LDL) cholesterol led to the identification of loss-of-function mutations in the gene coding for PCSK9, a key factor for lowering cholesterol (*5*) that is now the target of novel drug treatments (*6*).

Adaptation to diverse environments during human evolution has resulted in populations with phenotypes that are at the extremes of the trait distribution across all humans. Extreme phenotypes resulting from natural selection may in some cases be due to relatively few genetic variants with large effect. Integration of scans for natural selection, to narrow down candidate loci, with phenotype association studies can be particularly informative for identifying genetic variants associated with complex traits, even in studies with small sample sizes.

For example, this approach has been used to identify genetic loci influencing the extreme short stature trait in Central African hunter-gatherer populations, a trait thought to be an adaptation to a tropical environment (*7*, *8*). It has also been used to find loci influencing physiologic adaptation to high altitude in Tibet, South America, and Ethiopia (*9*). These studies

Departments of Genetics and Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: tishkoff@mail.med.upenn.edu



Global distribution of locally adaptive traits. Adaptation to diverse environments during human evolution has resulted in phenotypes that are at the extremes of the global distribution. Fumagalli *et al.* have integrated scans of natural selection and GWAS to identify genetic loci associated with adaptation to an Arctic environment.

ILLUSTRATION: A. CUADRA/*SCIENCE* AND MEAGAN RUBEL/UNIV. OF PENNSYLVANIA

show that smaller-scale studies of ethnically diverse populations living in a range of different environments and experiencing local adaptation are important and complementary to large genome-wide association studies (GWAS) of urban populations (see the figure).

The study of indigenous populations can be informative for several reasons. For example, although environments may vary considerably between populations, the environments within individual populations are often relatively homogeneous (similar diets, cultural habits, and environmental exposures). This can improve the power to identify genetic factors influencing complex phenotypes. In addition, increased levels of inbreeding in some small indigenous populations lead to regions of the genome that are identical within and among individuals and that can be useful for identifying rare variants influencing complex traits (*10*).

More fundamentally, investigation of such populations is crucial to further distribute the benefits of "precision medicine" and the associated development of targeted therapeutics. However, indigenous populations, particularly those living in rural areas with little access to health care, can be very challenging to study. Great care must be taken to avoid coercion and to be respectful of local practices, customs, and beliefs. This issue has been of particular concern to some Native American and Australian Aboriginal populations who may not want to participate in genomics research because of cultural beliefs. Some indigenous populations are also wary of genomics research owing to concerns about exploitation. There is valid concern about commercialization of information obtained from studies of indigenous populations. However, that concern must be balanced by the important need to develop therapeutics that will benefit all populations, not just those in developed countries. Thus, dialogue between researchers and indigenous communities is needed to ensure benefit sharing. It is critical to include diverse peoples in genomic studies to understand the phenotypic impact of the full range of human genetic variation and to ensure that all peoples benefit from this knowledge. ■

REFERENCES

1. A. R. Wood *et al.*, *Nat. Genet.* **46**, 1173 (2014).
2. A. E. Locke *et al.*, *Nature* **518**, 197 (2015).
3. M. Fumagalli *et al.*, *Science* **349**, 1343 (2015).
4. B. F. Voight *et al.*, *PLOS Genet.* **8**, e1002793 (2012).
5. J. C. Cohen, H. H. Hobbs, *Science* **340**, 689 (2013).
6. C. Sheridan, *Nat. Biotechnol.* **33**, 785 (2015).
7. J. P. Jarvis *et al.*, *PLOS Genet.* **8**, e1002641 (2012).
8. G. H. Perry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E3596 (2014).
9. L. B. Scheinfeldt, S. A. Tishkoff, *Genome Biol.* **11**, 133 (2010).
10. S. R. Browning, E. A. Thompson, *Genetics* **190**, 1521 (2012).

10.1126/science.aad0584

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

CANCER

Cancer therapies that are gone with the Wnt

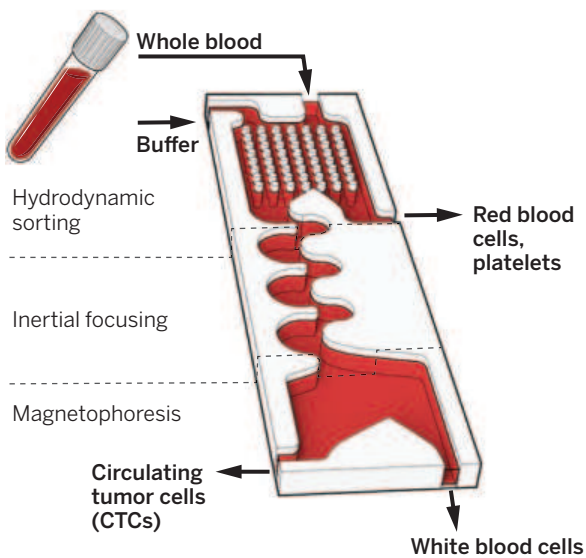
Wnt signaling in prostate cancer cells may contribute to escape from androgen receptor–targeted therapies

By **David M. Nanus** *and* **Paraskevi Giannakakou**

The isolation and molecular characterization of circulating tumor cells (CTCs) to guide cancer therapy is a particularly attractive approach for men with metastatic prostate cancer where the primary tumor may have been removed years earlier and the only indication of disease progression may be a rising amount of prostate-specific antigen (PSA) in the blood and a worsening bone scan (the skeleton is commonly the first area of metastasis for this cancer). Consequently, numerous technologies to capture, isolate, and study prostate cancer CTCs have been developed, each with its own limitations. Because of CTC heterogeneity, isolation approaches that rely on cell surface protein expression may not identify the prostate cancer cells undergoing epithelial-to-mesenchymal transition (which contributes to the migratory capacity and invasiveness of CTCs) (*1*). On page 1351 of this issue, Miyamoto *et al.* (*2*) report on the use of a microfluidic device that enables the capture of live CTCs amenable to single-cell molecular characterization (*3*). The study reveals a possible mechanism underlying the resistance of certain prostate cancer patients to androgen-targeted therapies.

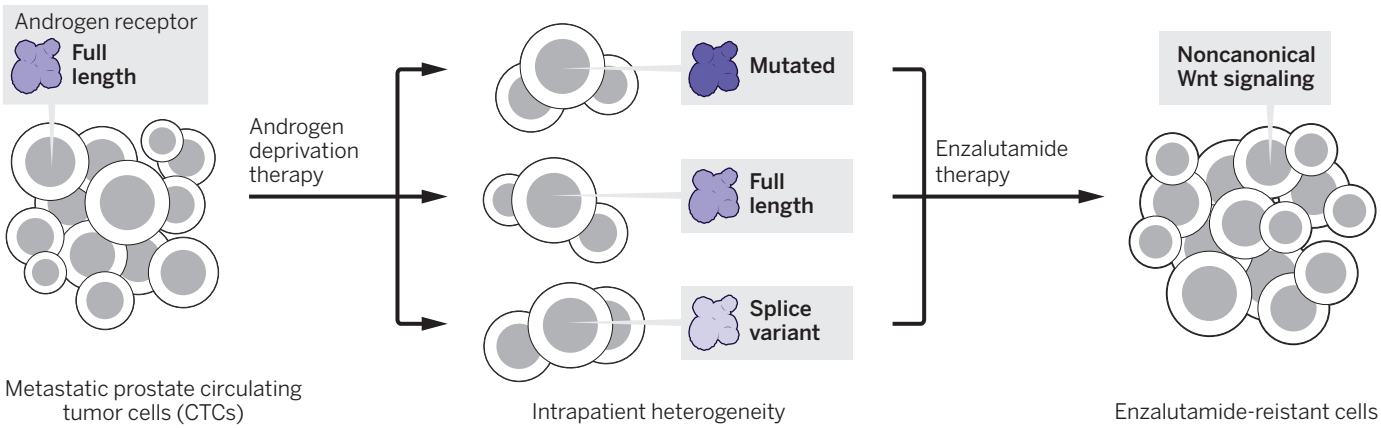
Prostate cancer growth is driven by male sex hormones called androgens, the most common of which is testosterone. Androgen deprivation therapy therefore remains the most effective treatment for patients with metastatic "castration-sensitive" prostate cancer. This is commonly accomplished through medical castration with drugs that inhibit testosterone production. For many men with "castration-resistant" prostate cancer, therapy frequently involves newer agents such as abiraterone (an androgen synthesis inhibitor) and enzalutamide [an androgen receptor (AR) antagonist or "anti-androgen"]. Retrospective studies examining prostate cancer CTCs that sought to decipher resistance to androgen signaling–targeted therapies (*4*, *5*) have directly led to larger prospective studies using multiple CTC isolation technologies to identify biomarkers of sensitivity and resistance to abiraterone and enzalutamide, including splice variants of AR-encoding messenger RNA (mRNA).

Miyamoto *et al.* used an antigen-agnostic microfluidic device called CTC-iChip (*3*) (see the first figure) to characterize gene expression in 77 single CTCs isolated from 13 patients with prostate cancer, of whom 11 had castration-resistant prostate cancer. The single-CTC RNA sequencing results provide several important observations that would not have been "resolved" in population-averaging RNA sequencing analyses. Single-CTC analysis showed that only one-sixth of the CTCs co-expressed more than one AR splice variant (AR-V), which is in sharp contrast to



CTC-iChip. Single CTCs are isolated from the peripheral blood of prostate cancer patients following enrichment with a microfluidic device.

Division of Hematology and Medical Oncology, Department of Medicine, and Meyer Cancer Center, Weill Cornell Medical College, New York, NY 10021, USA. E-mail: dnanus@med.cornell.edu; pag2015@med.cornell.edu



Heterogeneity and resistance. Following androgen deprivation therapy, metastatic prostate cancer cells undergo alterations in the androgen signaling axis (as determined by single-cell CTC analysis), including AR mutations and AR mRNA splice variants. Following antiandrogen treatment (enzalutamide), resistant cells show an increase in noncanonical Wnt signaling.

the common belief that several variants are co-expressed in biologic contexts (prostate cancer tumor tissue and/or CTCs) and that they may be competing with full-length AR (AR-FL) for dimerization, which is required for transcriptional activity (*4*, *6*). Furthermore, a recent study showed that AR-Vs can homodimerize as well as heterodimerize among themselves or with AR-FL, which suggests that inhibition of AR-V dimerization might be an important therapeutic strategy for the treatment of castration-resistant prostate cancer (*7*). Although the results reported by Miyamoto *et al.* require independent validation and expansion to a larger cohort of prostate cancer patients, if their findings hold true and only a small minority of CTCs co-express AR-Vs, then the above therapeutic strategies may need to be reconsidered.

Miyamoto *et al.* did not observe substantial enrichment in AR-V expression in patients on enzalutamide treatment relative to the enzalutamide-naïve patient cohort. This is in contrast to a study reporting that AR-V7 expression in patients' CTCs was associated with resistance to treatment with enzalutamide and abiraterone (*4*). Although the small number of patients in both studies could account for these discordant results, single-CTC analysis of AR-V expression might be advantageous in future studies as a means of resolving the complexity around the interplay between AR-FL and AR-V expression, with implications for clinical decision-making.

To maximize the biological insights that can be gained from single-CTC RNA sequencing analyses and to decipher mechanisms of treatment resistance, Miyamoto *et al.* performed a retrospective gene expression analysis by dividing the limited number of patients into two groups: enzalutamide-naïve and those receiving enzalutamide treatment. Analysis of cell signaling pathways identified substantial enrichment for noncanonical Wnt signaling in CTCs from the enzalutamide-resistant patients.

Signaling by the secreted protein Wnt plays a critical role in embryonic development and the maintenance of stem cell populations (*8*), and genetic aberrations in the Wnt signaling network have been broadly implicated in several types of cancer, including colorectal, lung, breast, and metastatic liver cancer. Noncanonical Wnt signaling is activated by a subset of Wnt ligands (such as Wnt5a and Wnt7b) and controls several downstream pathways critical for cell survival, proliferation, and motility.

Several studies have shown increased expression of Wnt5a and other noncanonical Wnt signaling components in prostate cancer, especially during the epithelial-to-mesenchymal transition (*9*, *10*). Likewise, Miyamoto *et al.* noted activation of noncanonical Wnt signaling in a small subset of CTCs (approximately 3%) from enzalutamide-resistant patients (see the second figure). Interestingly, the data from this patient cohort revealed an inverse correlation between activation of the glucocorticoid receptor signaling pathway and noncanonical Wnt signaling. The glucocorticoid receptor can become activated in drug-resistant prostate cancer. This receptor signaling pathway activates genes that promote cancer cell survival and growth. Along these lines, results from sequencing studies of castration-resistant prostate cancer specimens have identified the noncanonical Wnt pathway as one of the major signaling cascades with genomic alterations, whereas such alterations were virtually absent in hormone-naïve primary prostate cancer (*11*), in accordance with the findings of Miyamoto *et al.*

Although noncanonical Wnt signaling has been suggested to play an important role in prostate cancer recurrence and resistance to androgen deprivation therapy, it would be of interest to investigate its potential role in resistance to taxanes. These are widely used chemotherapeutic drugs that disrupt microtubules and inhibit cell division. Given the involvement of Wnt signaling in actin cytoskeleton remodeling, the cross-talk between the actin and microtubule cytoskeletons, and the use of microtubules by AR for its translocation into the nucleus (*5*, *12*), deciphering the potential effects of noncanonical Wnt signaling activation in prostate cancer on microtubule-dependent AR trafficking may have therapeutic implications for both androgen deprivation therapy and taxane resistance.

To stay ahead of an effective treatment's selective pressure on the genotype that results in a cancer cell's evolution to a resistant phenotype, it is imperative to molecularly interrogate tumor cells over time. As demonstrated by Miyamoto *et al.*, the application of CTC isolation technologies in prostate cancer, together with single-cell sequencing and CTC molecular characterization, can provide new information in real time. This information not only offers insight on resistance mechanisms but may guide the choice of the next therapy or lead to the earlier use of combination therapies. Although prospective clinical interrogations in larger patient populations are required in future studies, the study by Miyamoto *et al.* brings us one step closer to CTC-guided precision medicine. ■

REFERENCES

1. J. Li *et al.*, *Crit. Rev. Clin. Lab. Sci.* **52**, 191 (2015).
2. D. T. Miyamoto *et al.*, *Science* **349**, 1351 (2015).
3. E. Ozkumur *et al.*, *Sci. Transl. Med.* **5**, 179ra47 (2013).
4. E. S. Antonarakis *et al.*, *N. Engl. J. Med.* **371**, 1028 (2014).
5. M. S. Darshan *et al.*, *Cancer Res.* **71**, 6019 (2011).
6. D. Robinson *et al.*, *Cell* **161**, 1215 (2015).
7. D. Xu *et al.*, *Cancer Res.* **75**, 3663 (2015).
8. M. Katoh, M. Katoh, *Clin. Cancer Res.* **13**, 4042 (2007).
9. H. Yamamoto *et al.*, *Oncogene* **29**, 2036 (2010).
10. T. S. Gujral *et al.*, *Cell* **159**, 844 (2014).
11. N. N. Yokoyama et al. *Am. J. Clin. Exp. Urol.* **2**, 27 (2014).
12. M. Thadani-Mulero *et al.*, *Cancer Res.* **74**, 2270 (2014).

10.1126/science.aad2448

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

ORGANIC CHEMISTRY

Enchained by visible light–mediated photoredox catalysis

Radical-chain processes can dominate the kinetics of photogenerated radical catalysts

By **Markus D. Kärkäs, Bryan S. Matsuura, Corey R. J. Stephenson**

Free radicals are exploited in biology, often through highly controlled enzymatic reactions, to drive many reactions that would be difficult via nonradical routes that transfer two electrons (*1*). In synthetic chemistry, visible-light photoredox catalysis has emerged as an economical and environmentally benign route for promoting free radical transformations in the lab (*2*–*4*). Although the initial light-sensitization steps are well established (*5*), insufficient attention has been dedicated to essential mechanistic features of the closed catalytic cycle (*6*). Several reports have hypothesized that these photocatalyzed reactions are terminated through a closed catalytic cycle, which delivers the final product and regenerates the ground state of the photosensitizer (PS). However, Cismesia and Yoon (*6*) highlight that some of the mechanistic proposals may be incomplete and may involve radical chains.

Free radical reactions have traditionally been initiated in the lab with group-transfer reagents such as azobisisobutyronitrile (AIBN) or triethylborane and O_2. The radical formed from the substrate (substrate•) eventually reacts to form the desired product. The radical initiators are often used in substoichiometric quantities but are not catalysts. Rather, these reactions proceed through self-propagating radical-chain pathways.

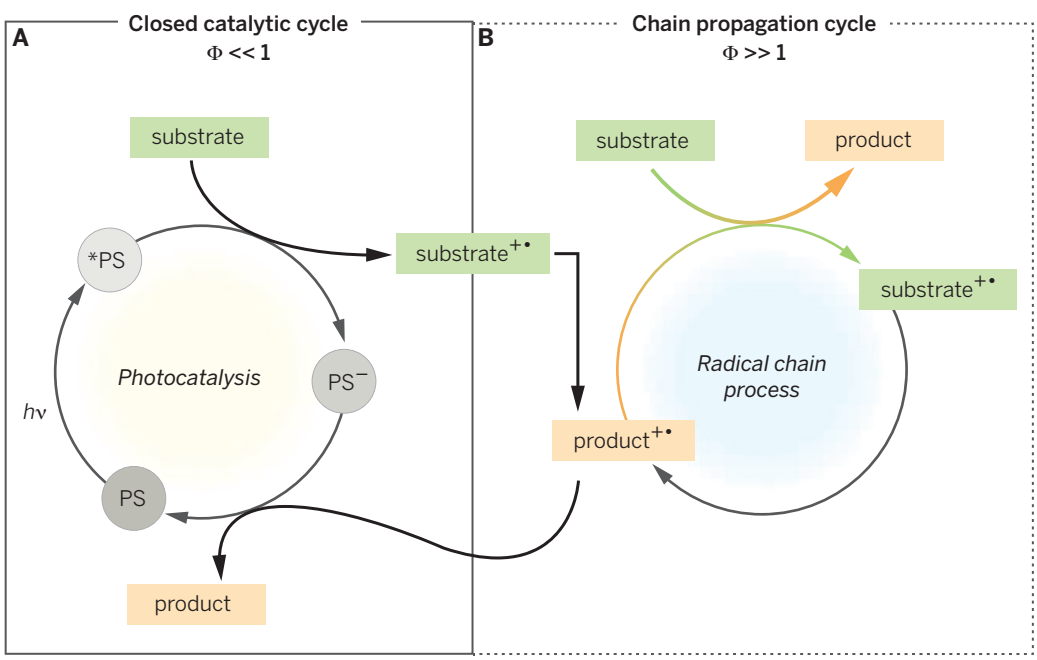
Photoredox catalysts generate radical intermediates through single-electron transfer and are initiated by light absorption by a photosensitizer (PS), which produces a long-lived excited triplet state (PS*) that can be either reductively quenched ($PS^{\bullet-}$) or oxidatively quenched ($PS^{\bullet+}$) by organic substrates. The overall result is the production of radical ions (substrate$^{\bullet+}$, substrate$^{\bullet-}$, or substrate•) that are subsequently converted to open-shell products (product$^{\bullet+}$, product$^{\bullet-}$, or product•) (see the figure, panel A).

> ***"A topic of debate concerning photoredox reactions is the extent to which chain processes are operative..."***

A topic of debate concerning photoredox reactions is the extent to which chain processes are operative, because group transfer generates the same intermediates as electron transfer. In chain processes, product formation would occur through a propagation step where the open-shell radical product reacts with another equivalent of the substrate (see the figure, panel B). Studying the reaction progress with alternating periods of illumination and darkness ("light-dark" experiments) has commonly been used to resolve these processes. That is, if the reaction stops in the dark cycle, it is assumed that chain processes did not occur or are relatively short-lived. However, the lifetimes for typical chain processes can be on the second or subsecond time scale (*7*), thus rendering it difficult to draw definitive conclusions from such experiments.

Cismesia and Yoon instead estimated the extent of chain processes in visible-light photoredox catalysis through a combination of quantum-yield (Φ) and luminescence-quenching measurements. In the quantum-yield measurements, the closed catalytic cycle can exhibit a maximum theoretical quantum yield of 1 (Φ = 1); every photon absorbed produces one molecule of product. A quantum yield of >>1 would suggest that chain processes are the major mechanistic pathway, whereas a quantum yield of <1 is inconclusive and consistent with both an efficient closed catalytic cycle and an inefficient chain process.

They studied three mechanistically distinct visible light–mediated reactions: The radical cation Diels-Alder cycloaddition (*8*); the radical anion [2+2] cycloaddition of enones (*9*); and the α-alkylation reaction of aldehydes, which proceeds via a neutral radical intermediate (*10*). Light-dark experiments revealed that product formation occurred only during intervals of constant irradiation (suggesting limited chain pro-



Radically different. Schematic depiction for photoredox catalysts of (**A**) a closed catalytic cycle and (**B**) a chain-propagation cycle. The approach of Cismesia and Yoon differentiates the two reaction mechanisms on the basis of quantum yield (Φ) and luminescence quenching. For the three reactions they studied, all proceeded through a chain-propagation cycle.

ILLUSTRATION: N. CARY/*SCIENCE*

Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: crjsteph@umich.edu

cesses), but for these three photocatalyzed reactions, quantum yields of >>1 were observed, indicative of radical-chain processes being dominant.

Care must be taken not to extrapolate these findings to all photoredox reactions. Instead, they are likely to be case-dependent (*11*, *12*). Mechanistic studies concerning photocatalyst quenching dynamics with luminescence quenching and transient absorption spectroscopy have played an important role in elucidating the elementary steps in catalyst-substrate interactions (*13*). Other mechanistic considerations, such as catalyst deactivation, have been studied with reaction-progress kinetic analysis (*14*). Collectively, these techniques provide the powerful tools necessary to probe the intricate mechanisms inherent in these catalytic systems. Further advancement of mechanistic understanding will have a tremendous impact on the development of novel transformations, the optimization of existing reactions, and the design of more effective photocatalysts. ■

"Collectively, these techniques provide the powerful tools necessary to probe the intricate mechanisms inherent in these catalytic systems."

REFERENCES AND NOTES

1. C. Chatgilialoglu, A. Studer, Eds., *Encyclopedia of Radicals in Chemistry, Biology and Materials* (Wiley, Weinheim, Germany, 2012).
2. D. M. Schultz, T. P. Yoon, *Science* **343**, 1239176 (2014).
3. J. J. Douglas, J. D. Nguyen, K. P. Cole, C. R. J. Stephenson, *Aldrichim. Acta* **47**, 15 (2014).
4. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322 (2013).
5. A. Juris *et al.*, *Coord. Chem. Rev.* **84**, 85 (1988).
6. M. A. Cismesia, T. P. Yoon, *Chem. Sci.* 10.1039/C5SC02185E (2015).
7. G. M. Burnett, H. W. Melville, *Chem. Rev.* **54**, 225 (1954).
8. S. Lin, M. A. Ischay, C. G. Fry, T. P. Yoon, *J. Am. Chem. Soc.* **133**, 19350 (2011).
9. M. A. Ischay, M. E. Anzovino, J. Du, T. P. Yoon, *J. Am. Chem. Soc.* **130**, 12886 (2008).
10. D. A. Nicewicz, D. W. C. MacMillan, *Science* **322**, 77 (2008).
11. H. Ismaili, S. P. Pitre, J. C. Scaiano, *Catal. Sci. Technol.* **3**, 935 (2013).
12. M. Majek *et al.*, *Beilstein J. Org. Chem.* **10**, 981 (2014).
13. N. A. Romero, D. A. Nicewicz, *J. Am. Chem. Soc.* **136**, 17024 (2014).
14. J. J. Devery III *et al.*, *Chem. Sci.* **6**, 537 (2015).

ACKNOWLEDGMENTS

M.D.K. gratefully acknowledges financial support from the Swedish Research Council (637-2013-7314) and the Royal Swedish Academy of Agriculture and Forestry (Kungliga Skogs- och Lantbruksakademien) for a postdoctoral fellowship.

10.1126/science.aad0193

ENERGY

King Coal and the queen of subsidies

The window for fossil fuel subsidy reform is closing fast

By **Ottmar Edenhofer**

Coal is the most important energy source for the Chinese economy (see the photo). Other rapidly growing economies in Asia and Africa also increasingly rely on coal to satisfy their growing appetite for energy. This renaissance of coal is expected to continue in the coming years (*1*) and is one of the reasons that global greenhouse gas (GHG) emissions are increasing despite the undisputed worldwide technological progress and expansion of renewable technologies (*2*). The implications for long-term GHG emissions are serious because, once installed, a coal power plant will emit for decades. Fossil fuel subsidies support investments in coal capacities around the globe and thereby threaten the achievement of climate change mitigation goals. Targeted reform of these subsidies could yield benefits for climate change mitigation as well as other development objectives.

The existing global energy infrastructure already commits 729 gigatons of CO_2 ($GtCO_2$) of future cumulative emissions over its lifetime. Aims to limit global temperature increase to 2°C allow for a total of 1000 $GtCO_2$ to be released into the atmosphere. If only one-third of currently planned coal capacity is installed successfully, an additional 113 $GtCO_2$ emissions are committed, nearly depleting the budget allowed by such mitigation targets (*3*).

Over the past year, many nations have made commitments to reduce their domestic GHG emissions. The U.S. government has announced plans to reduce emissions in the power sector by 32% below 2005 levels in 2030 through its Clean Power Plan (*4*). The Chinese government pledged that its emissions would peak in 2030 and has plans to increase the installed capacities of renewables and nuclear power substantially in the next decade (*5*). The Chinese emission target is sufficiently vague to prevent a precise evaluation of emission reductions. But even if China and the United States are successful in reducing their domestic emissions, worldwide emissions are expected to continue to rise. After all, a reduction in coal demand in one region reduces world market prices, incentivizing an increasing demand in other regions (*6*).

What explains this renaissance of coal? The short answer is the relative price of coal. The price of coal-based electricity generation remains much lower than that of renewable power when the costs of renewable intermittency are taken into account.

As a result of technological progress and economies of scale, the costs of generating electricity from wind and solar power have declined substantially. Wind generation now costs 70 US$ per megawatt-hour (MWh) on average, and geographically favorable sites can compete with the costs of coal-fired power (~50 US$/MWh). Solar photovoltaic projects have reached 80 US$/MWh and within a few years can also be expected to match the costs of coal generation (*7*, *8*). However, the costs of intermittency of wind and solar add an additional markup of about 30 US$/MWh (*9*) in cases where these technologies are deployed on a large scale as a result of increasing backup capacity requirements. Because of these additional costs, coal becomes more attractive for investors than renewable sources in many countries. In addition, coal is increasingly traded on the world market, dashing the hopes of many concerned with climate change that coal is only economically viable for a few countries with large domestic endowments (*1*).

At the same time, finance ministers around the globe subsidize fossil fuels, mostly by enabling the sale of these fuels on the domestic market below world market prices. In 2013, these pretax subsidies amounted to about 550 billion US$ worldwide (*10*). Energy subsidies are often believed to mainly support low-income households, but this belief is not well supported. Energy subsidies are typically captured by rich households in low-income countries and do little to support the poor. In all regions, the poorest 20% of the population received less than 8% of the benefits of the subsidies, whereas more than 40% of the subsidies were captured by the richest 20% (*11*).

Well-designed fossil fuel subsidy reform has considerable potential to raise the financial means necessary to reduce poverty. As a

Mercator Research Institute on Global Commons and Climate Change, Potsdam Institute for Climate Impact Research, Technische Universität Berlin, Germany. E-mail: ottmar.edenhofer@pik-potsdam.de

Coal renaissance. Rapidly growing economies, including China, increasingly rely on coal for cheap energy, jeopardizing efforts to reduce fossil fuel use worldwide. In this image, workers haul coal to barges in Fengjie for delivery to power plants downriver.

PHOTO: © BOB SACHA/CORBIS

recent study has shown (*12*), if current fossil fuel subsidies were to be redirected to investments in basic infrastructures over the next 15 years, substantial strides could be made in reducing poverty. This includes universal access to clean water in about 70 countries, to improved sanitation in about 60 countries, and to electricity in about 50 countries (out of roughly 80 countries that do not yet have universal access). Such investments would also increase the long-term growth prospects of poor economies.

The lion's share of pretax fossil fuel subsidies is targeted at oil consumption in the Middle East, North Africa, and Asia. At first glance, it would thus seem that subsidies are not instrumental in driving the renaissance of coal. However, this would be a premature conclusion: King Coal and the queen of subsidies are involved in a complex marriage, as shown by a recent International Monetary Fund (IMF) report on energy subsidies (*13*).

The IMF report suggests that fiscal spending is an incomplete and even misleading metric for fossil fuel subsidies: The nonpricing of adverse external effects such as GHG emissions, premature deaths through local air pollution, increased congestion, and other adverse side effects of vehicle use must also be taken into account. These subsidies also discourage investments in low-carbon alternatives such as energy efficiency, renewable energies, or natural gas infrastructure, which generate less or none of these externalities. In addition, the fiscal costs of subsidies must be financed by some combination of higher public debt, higher tax burdens, and crowding out of potentially productive public spending on health, education, and infrastructure.

A key insight from the IMF report is that posttax energy subsidies (which would include a price on the aforementioned externalities) are higher than the pretax subsidies by a factor of almost 10. The IMF calculates that the full social costs of fossil fuel consumption in 2013 amounted to 4.9 trillion US$ globally. A second major insight from their calculation is that coal receives about 60% of the total posttax subsidies. This implies that one ton of CO_2 receives, on average, more than 150 US$ in subsidies (with 32 $GtCO_2$ emitted by the global energy sector in 2013). The report convincingly shows that the mispricing of fossil fuels will contribute to an ongoing renaissance of coal over the coming decades.

The window of opportunity for successful price reform is rapidly closing. As many quickly growing countries continue to invest in coal-fired plants, they lock in carbon-intensive infrastructure, which substantially increases the costs of future emission reductions. Getting prices right before this infrastructure is built is essential. If the opportunity to correct the distortion in fossil fuel pricing is missed, climate policy is in peril.

The social costs of fossil fuel subsidies may not be obvious to the public and might even be masked for finance ministers. The upside of this debate is that adopting a more rational approach to fossil fuel pricing would increase overall welfare, provide fiscal gains for governments, and allow for new strategies to finance sustainable development that would particularly benefit the poor. These incentives arise from a purely self-interested national perspective, without the need to wait for a global climate agreement to come to fruition. At the same time, bold national actions to align fossil fuel prices with their true social costs could also remove important barriers for carbon pricing and hence become a major boost for international climate diplomacy. ■

REFERENCES

1. J. C. Steckel, O. Edenhofer, M. Jakob, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3775 (2015).
2. International Energy Agency (IEA), World Energy Outlook Special Report 2015: Energy and Climate Change (2015); www.iea.org/publications/freepublications/publication/weo-2015-special-report-energy-climate-change.html.
3. C. Shearer, N. Ghio, L. Myllyvirta, T. Nace, *Boom and Bust: Tracking the Global Coal Plant Pipeline* (Coalswarm, Sierra Club, 2015); http://action.sierraclub.org/site/DocServer/Coal_Tracker_report_final_3-9-15.pdf?docID=17381.
4. U.S. Environmental Protection Agency, Clean Power Plan for Existing Power Plants (2015); www2.epa.gov/cleanpowerplan/clean-power-plan-existing-power-plants.
5. White House, U.S.-China Joint Announcement on Climate Change (2014); www.whitehouse.gov/the-press-office/2014/11/11/us-china-joint-announcement-climate-change.
6. H. McJeon *et al.*, *Nature* **514**, 482 (2014).
7. IRENA, *Renewable Power Generation Costs in 2014* (International Renewable Energy Agency, Abu Dhabi, 2015).
8. Intergovernmental Panel on Climate Change, Working Group III, O. Edenhofer *et al.*, Eds., *Special Report on Renewable Energy Sources and Climate Change Mitigation* (Cambridge Univ. Press, Cambridge, 2011).
9. L. Hirth, F. Ueckerdt, O. Edenhofer, *Renew. Energy* **74**, 925 (2015).
10. IEA, World Energy Outlook 2014 (IEA, Paris, 2014).
11. F. J. Arze del Granado, D. Coady, R. Gillingham, *World Dev.* **40**, 2234 (2012).
12. M. Jakob, C. Chen, S. Fuss, A. Marxen, O. Edenhofer, *Nat. Climate Change* **5**, 709 (2015).
13. D. Coady, I. Parry, L. Sears, B. Shang, *How Large Are Global Energy Subsidies?* IMF working paper (2015); www.imf.org/external/pubs/ft/wp/2015/wp15105.pdf.

10.1126/science.aad0674

MINERAL CHEMISTRY

How minerals dissolve

Local surface processes drive calcite dissolution

By Mariëtte Wolthers[1,2]

During mineral growth and dissolution, material (in the form of atoms, molecules, complexes, or clusters) is transported through water to or from the mineral surface. Even in a solution that is at equilibrium with respect to a given mineral, there is transport to and from the surface. Although this equilibrium transport is in balance over large scales, the rates of this transport can vary locally depending on the energy landscape of the mineral surface (*1*). On page 1330 of this issue, Laanait *et al.* (*2*) present an elegant way of measuring and visualizing how dissolution rates vary across the calcite surface with time and with dissolution mechanism.

Calcite ($CaCO_3$) is the main long-term sink for carbon (*3*) and is one of the most abundant minerals in Earth's surface environment. It is therefore of prime importance in the global carbon cycle. It also helps to regulate the chemistry of aquatic environments (*4*), scavenges trace and heavy metals (*5*), and records paleoenvironmental conditions (*6*). One of the remaining challenges is the accurate prediction of how calcite dissolution rates vary and evolve (*1*). This is vital, for example, for preventing calcite dissolution in historic limestone and marble buildings and artifacts. This knowledge may also help to increase oil and gas yield from carbonate rock reservoirs.

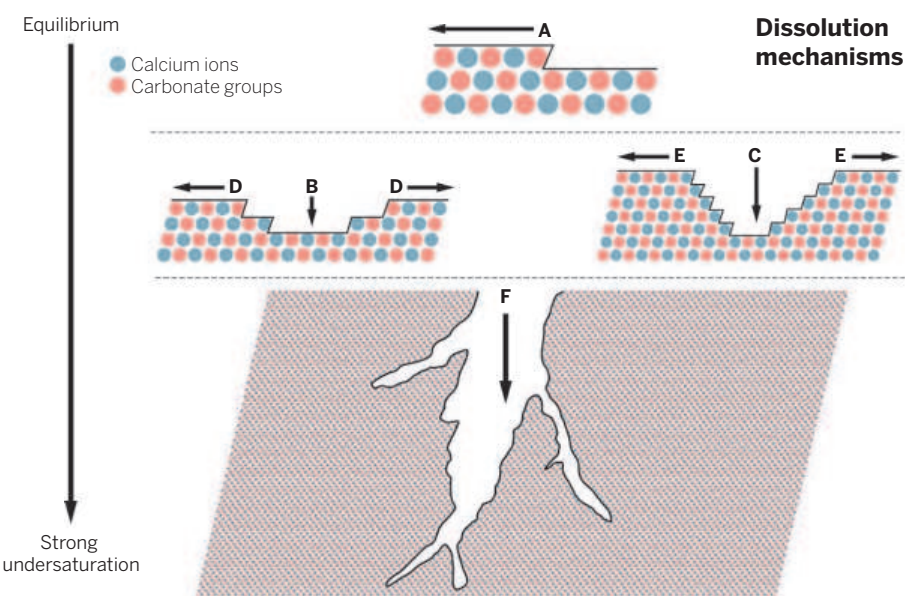
Following numerous macroscopic investigations of calcite dissolution with traditional methods (*4*), in the past decade research has zoomed in on the microscopic scale. In particular, atomic force microscopy (AFM) and interferometry have been instrumental for the direct observation of calcite dissolution mechanisms and the quantification of their kinetics (*5*). As in the current work (*2*), AFM and interferometry studies of calcite dissolution are performed on cleaved surfaces and have revealed a range of dissolution processes (see the figure). How far out of equilibrium a solution is will control which mechanism dominates calcite surface dissolution (*1*, *5*). Laanait *et al.* also report the formation of "wormholes," which occur in carbonate rocks upon CO_2 injection (*7*) but have not previously been shown using AFM.

Calcite dissolution studies by AFM have generally been performed in cells flushed with fluid of a constant composition. In contrast, Laanait *et al.* used the reaction of an intense x-ray beam with water close to the calcite surface. Radicals formed through this reaction drive fluid compositional changes that subsequently cause calcite dissolution. By varying x-ray beam intensity and exposure time, they could vary fluid composition and thereby scan a range of dissolution mechanisms and rates.

The lowest dissolution rates they report for etch-pit spreading are generally in good agreement with those reported using AFM (*5*), but the rates vary strongly and asymmetrically and can be almost five times as fast as those typically observed using AFM. These higher rates may be caused by the interaction of radicals in the fluid with the calcite surface or by energetically induced reactions caused by the x-ray beam. However, rate variability and oscillations between the fastest- and slowest-dissolving surface structures do not correlate with radical concentration (compare Fig. 4 with fig. S8). Variability in dissolution rates has been reported previously (*1*). Currently, distribution rate spectra are used instead of single-rate laws to simulate this variability. Laanait *et al.* now show quantitatively how calcite dissolution rate spectra evolve with surface structure, time, and dissolution mechanism.

The quantitative imaging technique that enabled revealing these large local variations in dissolution rate is known as particle image velocimetry (PIV). This technique has been used for a few decades (*8*) to determine the flow of suspended particles; Laanait *et al.* adapted it to map dissolution front velocities at the surface from surface x-ray microscopy. Combining PIV with time-resolved surface analysis techniques including AFM and interferometry will ultimately reveal the evolution of mineral surface dissolution rates over longer time scales and over a wide range of natural conditions. ■

[1]Department of Earth Sciences–Geochemistry, Faculty of Geosciences, Utrecht University, Netherlands. [2]Department of Chemistry, University College London, London, UK.
E-mail: m.wolthers@uu.nl



Calcite dissolution mechanisms. Calcite may dissolve through retreat of preexisting atomic steps at the calcite surface (**A**) (*9*), formation of new shallow (**B**) and deep (**C**) etch pits (*10*), widening of shallow etch pits (**D**) (*11*), and (**E**) trains of surface steps (stepwaves) that originate at deep etch pits (*1*, *12*). The number of etch pits formed depends on the number and type of imperfections in the surface. Shallow etch pits nucleate at surface defects and—if the solution is far enough out of equilibrium—also at defect-free sites (*9*). In contrast, deep etch pits form at crystal lattice imperfections referred to as screw dislocations (*1*). Laanait *et al.* also report "wormholing" in unstable reaction fronts (**F**).

ILLUSTRATION: A. CUADRA/*SCIENCE*

REFERENCES

1. C. Fischer *et al.*, *Appl. Geochem.* **43**, 132 (2014).
2. N. Laanait *et al.*, *Science* **349**, 1330 (2015).
3. D. Archer, E. Maier-Reimer, *Nature* **367**, 260 (1994).
4. J. W. Morse *et al.*, *Chem. Rev.* **107**, 342 (2007).
5. E. Ruiz-Agudo, C. V. Putnis, *Mineral. Mag.* **76**, 227 (2012).
6. E. A. Boyle, *Earth Planet. Sci. Lett.* **53**, 11 (1981).
7. L. Luquot, P. Gouze, *Chem. Geol.* **265**, 148 (2009).
8. C. E. Willert, M. Gharib, *Exp. Fluids* **10**, 181 (1991).
9. H. H. Teng, *Geochim. Cosmochim. Acta* **68**, 253 (2004).
10. Y. Liang *et al.*, *Geochim. Cosmochim. Acta* **60**, 4883 (1996).
11. S. L. S. Stipp *et al.*, *Geochim. Cosmochim. Acta* **58**, 3023 (1994).
12. A. Lüttge *et al.*, *Geochim. Cosmochim. Acta* **67**, 1099 (2003).

10.1126/science.aad0852

Caffeine, the circadian clock, and sleep

Why is caffeine intake at bedtime a sleep disrupter?

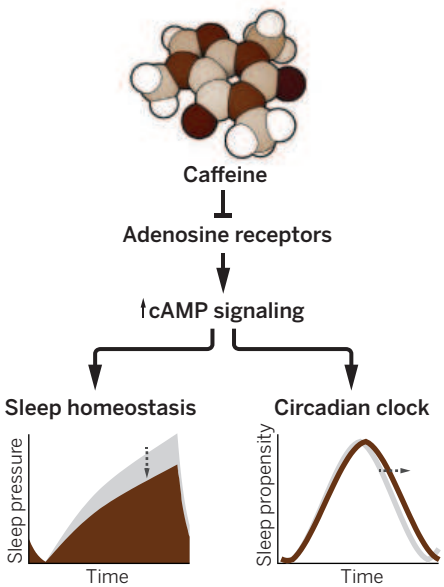
By **Hans Peter Landolt**

Caffeine wakes people up but also disrupts the quality of sleep. A new study by Burke *et al.* (*1*) reveals that consuming caffeine in the evening—the equivalent of a double espresso—delays the human endogenous circadian clock by antagonizing receptors for the endogenous sleep factor adenosine in the brain (see the figure). Mistimed caffeine consumption may contribute to the growing incidence of sleep problems in society.

Many people worldwide consume caffeine daily. Normal dietary consumption is sufficient to antagonize up to 50% of the inhibitory A_1 and the facilitatory A_{2A} adenosine receptors in the brain (*2*, *3*). This increases alertness and allays drowsiness and fatigue, but may also induce restlessness and prolong the time to fall asleep, enhance nighttime wakefulness, and reduce the depth of sleep (*4*).

Adequate sleep is required for good health and quality of life. The sleep-wake cycle is regulated by the fine-tuned interplay between homeostatic and circadian processes (*5*). Homeostatic sleep need accumulates during wakefulness and dissipates during sleep, whereas the circadian clock determines when sleep occurs. Slow-wave (or delta) neuronal activity (~0.75 to 4.5 Hz) recorded with an electroencephalogram (EEG) during deep sleep provides the most reliable biomarker of sleep need (*5*). Because caffeine attenuates sleep delta activity and blocks adenosine receptors, a role for adenosine and its receptors in sleep homeostasis has long been suggested (*4*). Burke *et al.* investigated whether caffeine also affects the human circadian clock. This is important because sleep and circadian systems are intimately linked at genetic, molecular, and behavioral levels.

Burke *et al.* used a highly sensitive protocol under strictly controlled conditions over a period of 49 days, and quantified the effects of 200 mg of caffeine on the timing of melatonin production in people when taken 3 hours before habitual bedtime in the evening. Melatonin is a hormone that in humans, entrains the circadian rhythm of many physiological processes, such as the timing of sleep, and is a reliable phase marker of the endogenous circadian pacemaker (*6*). Indeed, caffeine strongly and consistently delayed the melatonin rhythm by about 40 min, nearly half of the delay caused by bright light exposure at bedtime, a strong time cue for the circadian clock (*7*). But how does caffeine delay the circadian rhythm?



Double espresso effect. By blocking cerebral adenosine A_1 and A_{2A} receptors, caffeine increases intracellular cAMP signaling, attenuates the buildup of homeostatic sleep propensity during waking, and delays the circadian clock in vitro and in vivo.

The complex signaling cascade that regulates clock functions is expressed in nearly every cell of the body (*8*). In vitro, caffeine not only blocks adenosine receptors but also inhibits phosphodiesterase activity and activates ryanodine receptors. These actions increase cyclic adenosine monophosphate (cAMP)–dependent signaling and intracellular calcium release (*2*), both of which contribute to circadian timekeeping and resetting of the clock (*9*, *10*). Burke *et al.* measured circadian transcriptional rhythms in genetically engineered human cells expressing more than 10,000 proteins, including adenosine receptors, multiple phosphodiesterases, and ryanodine receptors. They confirmed that caffeine lengthens the circadian period and increases cAMP concentration. Although some open questions remain, their convergent pharmacological, genetic, and immunochemical data suggest an adenosine A_1 receptor–mediated, cAMP-dependent mechanism.

Caffeine-induced interference with the circadian clock may contribute to the high incidence of sleep problems in society and have a negative impact on brain functions that rely on undisturbed slow-wave sleep (*11*). Indeed, circadian rhythmicity modulates important functional characteristics of slow-wave sleep in humans (*12*). Yet, properly timed caffeine could alleviate jet lag and help patients with circadian sleep-wake disorders. Research on causal relationships among caffeine, circadian timekeeping, sleep, and health is warranted.

Apart from the circadian clock, cAMP signaling also plays an important role in sleep homeostasis and in the effects of caffeine on the consequences of sleep loss in animals (*13*, *14*). In humans, the physiological study of wakefulness and sleep is laborious, and the molecular mechanisms underlying sleep-wake regulation are difficult to elucidate. Given that cultured cells can display a sleep-like state (that is, neuronal firing activity reminiscent of sleep) (*15*), fundamental questions related to electrophysiological, genetic, and molecular features, as well as the pharmacology of sleep, can now be studied in human cells in vitro. Combined with physiological approaches, this opens up exciting new perspectives to examine the molecular bases of human sleep and to develop evidence-based therapeutic interventions for disturbed sleep in health and disease. ■

[1]*Institute of Pharmacology and Toxicology, University of Zürich, Zürich, Switzerland.* [2]*Zürich Center of Interdisciplinary Sleep Research, University of Zürich, Zürich, Switzerland. E-mail: landolt@pharma.uzh.ch*

REFERENCES

1. T. M. Burke *et al.*, *Sci. Transl. Med.* **305**, ra146 (2015).
2. B. B. Fredholm, J. F. Chen, S. A. Masino, J. M. Vaugeois, *Annu. Rev. Pharmacol. Toxicol.* **45**, 385 (2005).
3. D. Elmenhorst, P. T. Meyer, A. Matusch, O. H. Winz, A. Bauer, *J. Nucl. Med.* **53**, 1723 (2012).
4. H. P. Landolt *et al.*, *Neuropsychopharmacology* **29**, 1933 (2004).
5. P. Achermann, A. A. Borbély, in *Principles and Practice of Sleep Medicine*, M. H. Kryger, T. Roth, W. C. Dement, Eds. (Elsevier Saunders, St. Louis, MI, 2011), pp. 431–444.
6. A. J. Lewy, N. L. Cutler, R. L. Sack, *J. Biol. Rhythms* **14**, 227 (1999).
7. C. A. Czeisler *et al.*, *Science* **233**, 667 (1986).
8. A. Balsalobre, F. Damiola, U. Schibler, *Cell* **93**, 929 (1998).
9. J. S. O'Neill, E. S. Maywood, J. E. Chesham, J. S. Takahashi, M. H. Hastings, *Science* **320**, 949 (2008).
10. J. M. Ding *et al.*, *Nature* **394**, 381 (1998).
11. B. Rasch, J. Born, *Physiol. Rev.* **93**, 681 (2013).
12. A. S. Lazar et al., *Neuroimage* **116**, 123 (2015).
13. J. C. Hendricks *et al.*, *Nat. Neurosci.* **4**, 1108 (2001).
14. I. A. Alhaider et al., *Mol. Cell. Neurosci.* **46**, 742 (2011).
15. V. Hinard *et al.*, *J. Neurosci.* **32**, 12506 (2012).

10.1126/science.aad2958

CHEMICAL ENGINEERING

Cooling down ceramic fuel cells

Simplified processing and improved performance may lead to low-temperature fuel cells

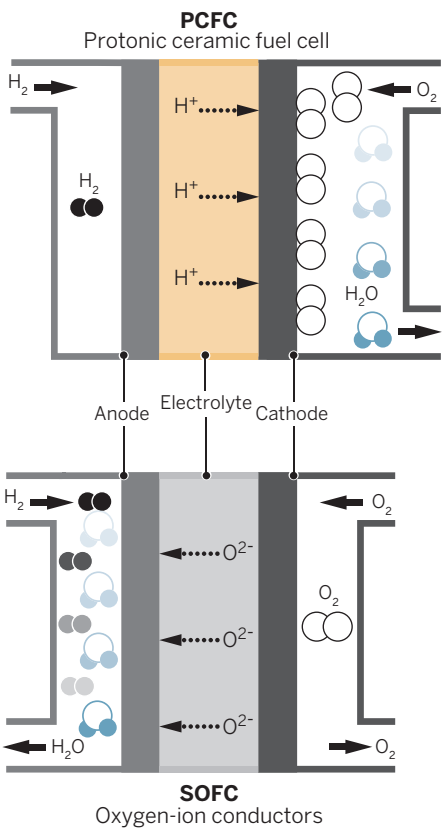
By **Raymond J. Gorte**

Ceramic fuel cells capable of achieving high power densities are based on oxygen-ion conductors that operate at high temperatures. The development of ceramic fuel cells that exhibit practical power densities at lower temperatures, with proton-conducting electrolytes, has been a long-standing dream, whose realization could lead to large-scale implementation of fuel cells. On page 1321 of this issue Duan *et al.* (*1*) report on three important contributions in the development of protonic ceramic fuel cells (PCFCs): the demonstration of the effective use of sintering aids to make difficult-to-prepare electrolytes that exhibit high protonic conductivities; the design of a new cathode material for PCFCs; and the development of a scalable fabrication process for cell production. Impressive performance was obtained with small-scale PCFCs at moderate temperatures. Taken together with other reports of high performance in PCFCs (*2*), practical ceramic fuel cells could be within reach.

Although the intrinsic conductivities of ceramic proton conductors have long been known to be higher than that of traditional oxygen-ion conductors, it has been difficult to take advantage of this in devices. The conductivities of the most promising proton-conducting ceramic, doped $BaZrO_3$, are usually limited by barriers at grain boundaries and are therefore sensitive to processing conditions. Duan *et al.* demonstrate that this problem can be solved with simple and inexpensive sintering aids. Improved cathode materials allow lower operating temperatures that avoid the materials-stability and corrosion problems that occur in traditional solid oxide fuel cells (SOFCs) (*3*). Operation above 700°C requires that the membrane-electrode assembly be made from special alloys that can withstand the extremely corrosive environment. High temperatures accelerate the coarsening of nanostructured electrodes. Also, startup from ambient conditions is simplified when the fuel cell generates power at low temperatures.

Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: gorte@seas.upenn.edu

PCFCs will likely find their most exciting applications with H_2 as the fuel, in devices where proton-exchange membrane fuel cells (PEMFCs) are presently being considered. The electrodes in PEMFCs almost always contain precious metals, whereas none is required by PCFCs. Also, PCFCs can operate at just the right temperatures. With PEMFCs, the electrolytes must be wet, limiting their maximum operating temperature to near the boiling point of water. Under these conditions, the precious-metal electrodes are highly susceptible to poisoning by impurities in the H_2, especially CO. Also, rejection of waste heat is difficult when operating near the ambient temperature. Indeed, much effort has gone into searching for proton-conducting electrolytes capable of operation up to 500°C (*4*).



Fuel cell operation. The generation of water in the anode compartment of traditional SOFCs results in the dilution of the H_2 fuel by H_2O, which can cause oxidation of the anode that can in turn cause cracking of the electrolyte. In PCFCs, H_2O is formed at the cathode, where it does not appreciably affect the O_2 concentration.

In addition to the lower operating temperatures compared to conventional SOFCs, there is no dilution of the H_2 by H_2O within the anode compartment of PCFCs (see the figure). Avoiding H_2 dilution with water improves performance at higher fuel conversions and averts the serious problem that Ni-based anodes can be oxidized to NiO in traditional SOFCs. Oxidation of Ni anodes in SOFCs is potentially catastrophic because it can cause cell fracture due to the expansion that occurs upon NiO formation.

For larger-scale fuel cells operating on natural gas, conventional SOFCs will likely still be preferred. Although H_2 can be generated internally from natural gas using the methane steam reforming (MSR) reaction, $H_2O + CH_4 \rightarrow CO + 3H_2$, temperatures greater than 500°C are required. Furthermore, system efficiency is increased when the waste heat generated by the fuel cell is consumed in driving the strongly endothermic MSR reaction. This is more easily accomplished in a conventional SOFC operating at higher temperatures (*5*). The MSR reaction can also be used for cooling to control cell temperature.

The demonstration of high protonic fluxes at just the right temperature also opens the exciting possibility of using PCFCs for producing H_2 for reactions. Friebe *et al.* (*6*) used a short-circuited PEMFC to separate H_2 from a mixture of gases, transporting the H_2 from one side of the cell to the other. However, higher PCFC operating temperatures greatly expand the number of possible reactions that can be done in such a membrane. For example, H_2 is industrially generated together with CO by the MSR reaction. The temperature range of 350° to 500°C offered by PCFCs is nearly ideal for producing extra H_2 from CO by the water–gas shift (WGS) reaction ($H_2O + CO \rightarrow H_2 + CO_2$). Removing H_2 as it forms theoretically allows the WGS reaction to be carried out to 100% conversion, so that both H_2 and CO are utilized.

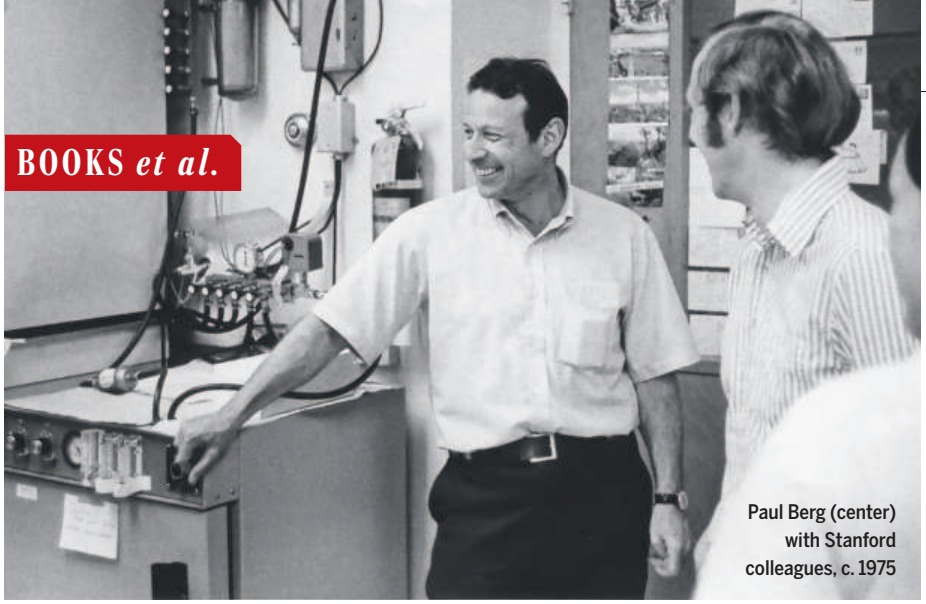
Much work still remains to take fuel cells from demonstration units to wide-scale, commercial reality. However, advances like those demonstrated by Duan *et al.* bring us one step closer to having practical devices for real applications by providing high performance at just the right temperatures, and using commercially viable fabrication procedures. ■

REFERENCES

1. C. Duan *et al.*, *Science* **349**, 1321 (2015).
2. J. Kim *et al.*, *ChemSusChem* **7**, 2811 (2014).
3. E. D. Wachsman, K. T. Lee, *Science* **334**, 935 (2011).
4. S. M. Haile, D. A. Boysen, C. R. Chisholm, R. B. Merle, *Nature* **410**, 910 (2001).
5. S. McIntosh, R. J. Gorte, *Chem. Rev.* **104**, 4845 (2004).
6. S. Freibe *et al.*, *Angew. Chem. Int. Ed.* **54**, 7790 (2015).

10.1126/science.aad0432

ILLUSTRATION: K. SUTLIFF/*SCIENCE*



Paul Berg (center) with Stanford colleagues, c. 1975

HISTORY OF SCIENCE

The cloning controversy

Lessons from the recombinant DNA revolution of the 1970s

By **Sally Smith Hughes**

The Recombinant University
Doogab Yi
University of Chicago Press, 2015. 318 pp.

A Biography of Paul Berg
Errol C. Friedberg
World Scientific, 2014. 406 pp.

The contemporary scientific community largely takes for granted the safety of recombinant DNA technology and its utility for practical application. It was not always so. The two books under review probe the science, politics, legal parameters, and practical repercussions of the so-called recombinant DNA revolution of the 1970s. They chronicle in detail the breakthroughs and contentions accompanying the invention of recombinant technology and the raucous controversies over its potential safety. Both books also tackle the resulting shift in ethical, aspirational, and legal landscapes in biomedical research as molecular scientists rushed to form industrial ties, universities accelerated patenting and licensing activities, and entrepreneurs created biotechnology companies.

This dramatic and convoluted history, which the authors pull together from an extensive body of preexisting written and oral sources, holds interest in its own right. But because similar issues find resonance and relevance today in current deliberations over embryonic and fetal stem cell research, appropriate guidelines for practicing CRISPR genome editing, and proper limits to patenting public knowledge, the reader should find it instructive to revisit this formative period of biological history and perhaps even garner lessons from it.

Doogab Yi, a Princeton-trained science historian at Seoul National University, and Errol Friedberg, an emeritus professor of pathology at the University of Texas Southwestern Medical Center and expert on DNA repair, focus on the research and researchers of Stanford Medical School's Department of Biochemistry. Here, in 1972, Paul Berg and his colleagues succeeded in covalently joining DNA molecules from different species, creating recombined or "recombinant" DNA. The technique was complicated, and required enzymes not readily available at the time. Berg's initial goal, to study the function of introduced DNA in foreign host cells, would eventually be overshadowed when researchers realized that the technique could be used to clone (replicate) DNA.

Although Friedberg and Yi examine the same department at the same time period, they have different objectives and take the genesis and consequences of recombinant DNA technology in divergent directions.

Friedberg's book is a straightforward biography of the multifaceted Paul Berg: biochemist, Nobel laureate, and statesman of science. Friedberg follows Berg from birth to his retirement years but highlights Berg's role in the recombinant DNA controversy.

In 1973, Stanley N. Cohen, a Stanford molecular geneticist, and Herbert W. Boyer, a University of California, San Francisco, biochemist, published a relatively simple and efficient technique for joining and cloning genes. The Cohen-Boyer method became the method of choice for genetic engineering in molecular biology. The newfound ease with which genes could be engineered to human expectations engendered widespread fears. Segments of the scientific community and public grew concerned that possibly dangerous life forms might threaten human health and upset the balance of nature. Scientists, with Berg in the lead as organizer and moral force, called for a temporary moratorium on recombinant DNA research until federal safety containment guidelines could be formulated within which the technology could be practiced and, importantly, the promising research advanced.

The fact that the book is aimed "primarily at scientists and those in training" (cover blurb) may explain why Friedberg, his molecular biology background much in evidence, devotes multiple chapters to recombinant science, largely that of Berg and his departmental colleagues, but also briefly Cohen and Boyer's work with a plasmid transfer system and a key restriction enzyme (EcoR1). Reflecting a scientist's concern for priority, he attempts to unravel competing scientific claims, particularly regarding the sticky question of how much Cohen benefited (conceptually and experimentally) from Stanford Biochemistry. Friedberg ultimately determines that the Cohen-Boyer method indeed built upon "seminal ideas and/or experiments" from Stanford and elsewhere (*1*, *2*).

In 1980, Berg received the Nobel Prize in Chemistry for "his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA." Cohen and Boyer were not mentioned. Friedberg concludes after reviewing relevant documents that "there is little question that Stanley Cohen and Herb Boyer merited recognition for their contribution to advancing the technology of gene cloning."

Yi's *The Recombinant University* is a very different book. Although he does not specify his intended audience, science historians and sociologists appear to be the targeted readership, given the book's objectives and tone. However, scientists willing to condone the sprinklings of academic jargon would also benefit from reading this heavily researched and highly contextual history of a seminal period in postwar American biomedical science. Despite his tight focus on Stanford Biochemistry, Yi's goal is sweeping: "My close examination of changing scientific

The reviewer is historian of science emerita at the Oral History Center, University of California, Berkeley, Berkeley, CA 94720, USA. E-mail: sshughes@berkeley.edu

PHOTO: FROM THE PERSONAL COLLECTION OF PAUL BERG

agendas, legal practices, and moral assumptions about commercialization in the [San Francisco] Bay Area academic community intends to tell a much broader story of the reconfiguration of both academic institutions and commercial enterprise in biomedical research." He adeptly shows how changes in American society from the 1960s on—notably, the demand for medical relevance in biomedical research; the Bayh-Dole Act of 1980, which gave academic institutions incentives to patent government-supported research results; and the Supreme Court's Chakrabarty decision, which enabled the patenting of human-made living organisms—facilitated the arrival of "the Recombinant University" with its enhanced interest in patenting and licensing and myriad ties with industry.

Yi's analysis shows how Stanford Biochemistry, an original citadel of basic science, adapted to accommodate the practical, financial, and legal vistas stemming from recombinant DNA technology. Even the department's two Nobel laureates, Arthur Kornberg and Paul Berg, despite their wariness about the commercialization of American biomedicine, fostered the department's industrial affiliates program and the biotechnology company DNAX, both founded in 1980.

But Yi's preoccupation with Stanford Biochemistry leaves this reviewer wondering: In the Bay Area's highly mobile and interactive scientific community, was recombinant DNA technology really overwhelmingly Stanford's creation? Only in passing does Yi address the significant research of Herbert Boyer's UCSF laboratory: its contribution and worldwide distribution of EcoR1, the restriction enzyme critical to the early performance of recombinant DNA technology, and its pioneering use of synthetic DNA in constructing genetic material. Where, also, is any mention of the role of chemically synthesized DNA as developed initially by Gobind Khorana of MIT and later used in Genentech's synthesis of the human insulin gene in 1978? Omissions such as these point to the need for further comparative historical work on the contributions of other institutions and departments at the forefront of genetic engineering in the 1970s.

Yet, Yi's broadly conceived and deeply interpreted analysis of a significant department at a major juncture in biomedical research contributes to a fuller understanding of the institutional, legal, and cultural transformations in American bioscience and society occurring at the time.

REFERENCES AND NOTES

1. For Cohen's adamant disagreement, see http://digitalassets.lib.berkeley.edu/roho/ucb/text/cohen_stanley.pdf.
2. For Berg's viewpoint, see http://content.cdlib.org/ark:/13030/kt1c6001df.

10.1126/science.aac9095

PHOTO: SVETIKD/ISTOCKPHOTO.COM

PROFESSIONAL DEVELOPMENT

Making contacts count

A scientist's guide to building professional networks

By **Monika S. Magoń**

The success of any scientist or engineer is the result not only of hard work and technical excellence, but also of a combination of creativity, problem-solving abilities, and, perhaps most important, interactions with other people. In *Networking for Nerds*, Alaina G. Levine offers advice on how to define, maintain, advance, and communicate your "brand"—what makes you unique and valuable as a scientist— to potential employers and collaborators.

Levine begins by dispelling common myths about networking, including the ideas that it's a waste of time, a sleazy salesman's tactic, and that only job seekers need to network. She highlights how networking can open otherwise hidden opportunities, including ones that might not yet exist, and argues that it also benefits the entire science and engineering community ("you cannot have innovation without regular influx of a diversity of ideas"). Levine goes on to show readers how to think about and communicate the value of their particular skills and experience, what she calls a "brand statement." Keep it simple and short, she advises: "a brand statement is not a thesis. It is not meant to take three hours to deliver."

In Chapter 4, Levine walks the reader through the steps needed to establish, maintain, and grow a professional network, emphasizing things like the importance of asking questions ("successful people remain successful by being inquisitive") and the value of a good mentor ("a mentor has access to networks you don't have access to"). In later chapters, the reader learns how to make any situation a networking opportunity, as well as how to employ social media to increase one's professional visibility.

The best advice comes from the practical pointers and real-life examples sprinkled throughout the book. These tips and vignettes help the reader think about how to incorporate similar strategies into one's own networking activities and to visualize the potential benefits that will be derived from these efforts.

Networking for Nerds serves as a guide for developing skills that are almost never taught in a formal way to scientists and engineers. This excellent resource should be on the reading list of those in the early stages of their careers, as well as those who may be contemplating a career change or hoping to secure a promotion.

10.1126/science.aad0594

Networking for Nerds
Find, Access and Land Hidden Game-Changing Career Opportunities Everywhere
Alaina G. Levine
Wiley Blackwell, 2015. 225 pp..

The reviewer is in the Biotechnology and Biological Sciences Research Council London Interdisciplinary Doctoral Programme at the University College London, London, UK. E-mail: m.magon.12@ucl.ac.uk

"When networking…you should be listening most of the time and talking very little," advises Alaina Levine. "But when you do talk, ask questions and insert information about your own experience that's relevant."

LETTERS

Edited by **Jennifer Sills**

Hunted predators: Charisma confounds

IN THEIR PERSPECTIVE "When the hunter becomes the hunted" (19 June, p. 1312), R. Woodroffe and S. M. Redpath rightly pointed out that a social consensus, in addition to scientific evidence, is critical to effectively control species deemed overabundant, such as some predators, because social conflicts over the need for control undermine management decisions. We stress here that social acceptability is not a sufficient criterion to prevent misguided lethal management of wildlife pests, as acceptability subjectively depends on the target species' charisma.

The black-footed ferret is an unintended victim of pest management.

Large predators exert fascination on the general public; hence, their control is usually unpopular [e.g., (*1*)], irrespective of the ecological desirability of management. By contrast, rodents and lagomorphs do not enjoy the same public appeal as predators. They are widely controlled on vast spatial scales, often with public funds, where they are perceived as creating damage to crops or forage used by livestock. Yet most of these persecuted herbivores play key ecological roles and are vital prey for emblematic and threatened predators (*2*, *3*). Here, the broad social acceptability of their control, due to their lack of charisma or low public visibility, may hinder evidence-based management decisions.

For example, the European hamster was driven from pest to red-listed critically endangered species by poisoning; the ruthless control of prairie dogs contributed to the critically endangered status of the black-footed ferret (*3*); and plateau pikas and voles have been poisoned over thousands of kilometers in China and Europe, respectively (*3*, *4*), with widespread unintended secondary poisoning of their birds of prey and mammalian predators (*5*, *6*). A focus on large predator control as a reference for wildlife management conflicts may overlook wider issues involving fauna that have keystone ecosystem roles but no public appeal, and where social consensus leads to misguided management. In the case of noncharismatic small herbivores, as long as public society awareness about their ecological benefits is not increased, their population control will continue to be widely accepted and possibly used without rigorous scientific evidence, thus jeopardizing biodiversity conservation.

Miguel Delibes-Mateos,[1,2]* François Mougeot,[3] Beatriz Arroyo,[3] Xavier Lambin[4]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado, University of Porto, 4485-661, Vairão, Portugal. [2]Instituto de Estudios Sociales Avanzados (IESA-CSIC), 14004, Córdoba, Spain. [3]Instituto de Investigación en Recursos Cinegéticos (IREC, CSIC-UCLM-JCCM), 13005, Ciudad Real, Spain. [4]Institute of Biological and Environmental Sciences, Zoology Building, Aberdeen AB24 2TZ, Scotland, UK.

*Corresponding author.
E-mail: mdelibesmateos@gmail.com

REFERENCES

1. M. Campbell, B. L. Lancaster, *Soc. Anim.* **18**, 40 (2010).
2. T. Cornulier *et al.*, *Science* **340**, 63 (2013).
3. M. Delibes-Mateos *et al. Biol. Conserv.* **144**, 1335 (2011).
4. J. J. Luque-Larena *et al. Basic Appl. Ecol.* **14**, 432 (2013).
5 M. Coeurdassier *et al., Conserv. Biol.* **28**, 315 (2014).
6. I. S. Sánchez-Barbudo *et al., Sci. Total Environ.* **420**, 260 (2012).

Hunted predators: Intrinsic value

IN THEIR PERSPECTIVE "When the hunter becomes the hunted" (19 June, p. 1312), R. Woodroffe and S. M. Redpath summarize issues associated with lethal "predator control"—the killing of carnivores to benefit human industries (e.g., livestock production). We applaud their careful accounting of the ecological and economic trade-offs that accompany such actions, but we take issue with their assertion that "[p]ragmatic conservationists have long recognized that allowing some predator control—whether or not it achieves its stated aims—can help to build tolerance...."

This claim is problematic for three reasons: First, existing research indicates that lethal control is often ineffective for increasing tolerance for wildlife (*1*, *2*). Second, as their own review makes clear, predator control often results in unforeseen ecological consequences (e.g., loss of ecosystem services). It is not only pragmatic for conservationists to oppose actions that may degrade ecosystem services and are ineffective for increasing tolerance, it is socially responsible. Finally, their claim relies on an ethical premise that few would accept: in essence, that it is acceptable to promote the killing of an organism as a means of reducing antipathy toward it. In fact, a recent study indicates most people believe that wildlife possess "intrinsic value," which suggests that wildlife should be treated with regard for their own welfare, not just their utility (or lack thereof) to humans (*3*).

From an ethical perspective, treating wildlife with regard to their own welfare would require shifting the burden of proof such that those who advocate the use of lethal methods would be forced to provide strong arguments as to why such means were justified. That we still manage wildlife in a manner that so easily dismisses the

PHOTO: USFWS MOUNTAIN-PRAIRIE/FLICKR

interests of wild animals and the preferences of the broader public suggests that the institution of wildlife management needs broader reforms than the additional stakeholder processes that Woodroffe and Redpath recommend.

Jeremy T. Bruskotter,[1]* Michael Paul Nelson,[2] John A. Vucetich[3]

[1]School of Environment and Natural Resources, The Ohio State University, Columbus, OH 43210, USA. [2]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA. [3]School of Forest Resources and Environmental Sciences, Michigan Technological University, Houghton, MI 49931, USA.

*Corresponding author. E-mail: bruskotter.9@osu.edu

REFERENCES

1. C. Browne-Nuñez, A. Treves, D. MacFarland, Z. Voyles, C. Turng, *Biol. Conserv.*, in press (2015); http://faculty.nelson.wisc.edu/treves/wolves/wolfpubs/Browne-Nunez_2015.pdf.
2. A. Treves, J. T. Bruskotter, *Science* **344**, 476 (2014).
3. J. A. Vucetich, J. T. Bruskotter, M. P. Nelson, *Conserv. Biol.* **29**, 321 (2015).

Response

OUR PERSPECTIVE highlighted the need to base predator management on scientific evidence that is not only technically sound, but also accepted by stakeholders. M. Delibes-Mateos *et al.* argue that social acceptability alone is insufficient to justify control efforts. We entirely agree. As we explained in our Perspective, controlling populations of predators (and, by extension, other species with strong ecological interactions) can have unintended ecological consequences. Hence, decisions about control efforts should be informed by evidence of the likely impacts.

J. T. Bruskotter *et al.* raise concerns about our statement that "[p]ragmatic conservationists have long recognized that allowing some predator control—whether or not it achieves its stated aims—can help to build tolerance among land managers who might otherwise block conservation efforts." They do not acknowledge the caveat that followed it: "Unfortunately, such compromise is not always effective." Far from advocating ineffective predator control, we highlighted deficiencies in the evidence underlying some control programs and proposed an alternative approach.

Bruskotter *et al.* also suggest that decision-making should incorporate the ethical perspective of the general public, such that "those who advocate the use of lethal methods...be forced to provide strong arguments as to why such means [are] justified." We agree that ethical considerations should inform decision-making, but we question the practicality of their suggestion. Decisions about carnivore management are made not just by "the institution of wildlife management" but also by private individuals, often acting illegally, potentially influencing predator abundance over large areas (*1*, *2*). Moreover, as we indicated in our Perspective, the "strong arguments" advanced by some stakeholder groups for or against predator control may not be accepted by others who hold different values. For example, evidence that localized killing of badgers increases disease risks to cattle (*3*) appears not to have deterred farmers from pursuing this practice illegally (*4*), undermining disease control efforts (*5*). Such observations underpin our call for engagement of multiple stakeholders to develop an agreed-upon evidence base, maximizing opportunities for both institutions and individuals to make decisions based on scientifically robust information.

Rosie Woodroffe[1]* and Stephen M. Redpath[2]

[1]Institute of Zoology, Zoological Society of London, London NW1 4RY, UK. [2]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK.

*Corresponding author. E-mail: rosie.woodroffe@ioz.ac.uk

REFERENCES

1. R. Woodroffe, L. G. Frank, *Anim. Conserv.* **8**, 91 (2005).
2. B. Etheridge, R. W. Summers, R. Green, *J. Appl. Ecol.* **34**, 1081 (1997).
3. C. A. Donnelly *et al.*, *Nature* **426**, 834 (2003).
4. P. Cross, F. A. V. St John, S. Khan, A. Petroczi, *PLOS ONE* **8**, 6 (2013).
5. D. M. Wright *et al.*, *Sci. Rep.* **5**, 13062 (2015).

ERRATA

Erratum for the Review "Plant and animal sensors of conserved microbial signatures" by P. C. Ronald and B. Beutler, *Science* **349**, aad3218 (2015). Published online 11 September 2015; 10.1126/science.aad3218

Erratum for the Report "The psychological consequences of money" by K. D. Vohs *et al.*, *Science* **349**, aac9679 (2015). Published online 24 July 2015; 10.1126/science.aac9679

Erratum for the Report "Reaction of O_2 with subsurface oxygen vacancies on TiO_2 anatase (101)" by M. Setvín *et al.*, *Science* **349**, aac9659 (2015). Published online 17 July 2015; 10.1126/science.aac9659

Erratum for the Report "The fastest unbound star in our Galaxy ejected by a thermonuclear supernova" by S. Geier *et al.*, *Science* **349**, aac9469 (2015). Published online 10 July 2015; 10.1126/science.aac9469

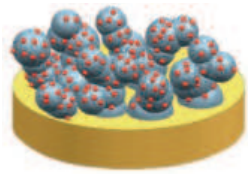
Erratum for the Report "On the origin of near-infrared extragalactic background light anisotropy" by M. Zemcov *et al.*, *Science* **349**, aac9468 (2015). Published online 10 July 2015; 10.1126/science.aac9468

Erratum for the Report "Morality in everyday life" by W. Hofmann *et al.*, *Science* **348**, aac5401 (2015). Published online 15 May 2015; 10.1126/science.aac5401

RESEARCH

Building a better proton-conducting ceramic fuel cell

Duan et al., *p. 1321*



IN *SCIENCE* JOURNALS

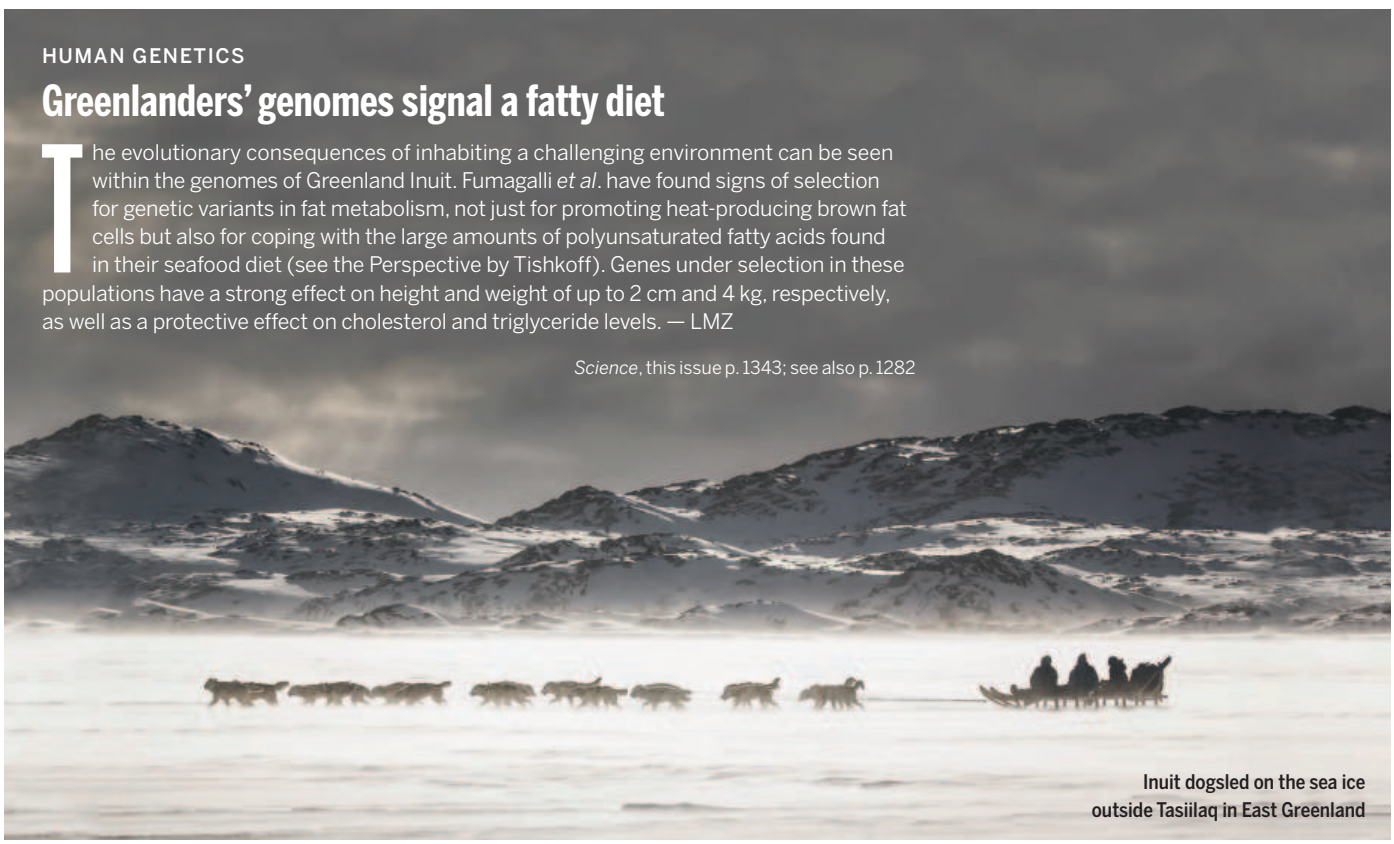
Edited by **Caroline Ash**

HUMAN GENETICS

Greenlanders' genomes signal a fatty diet

The evolutionary consequences of inhabiting a challenging environment can be seen within the genomes of Greenland Inuit. Fumagalli *et al*. have found signs of selection for genetic variants in fat metabolism, not just for promoting heat-producing brown fat cells but also for coping with the large amounts of polyunsaturated fatty acids found in their seafood diet (see the Perspective by Tishkoff). Genes under selection in these populations have a strong effect on height and weight of up to 2 cm and 4 kg, respectively, as well as a protective effect on cholesterol and triglyceride levels. — LMZ

Science, this issue p. 1343; see also p. 1282



Inuit dogsled on the sea ice outside Tasiilaq in East Greenland

PHOTO: MADS PHIL AND VISIT GREENLAND

VACCINES

Flu vaccine candidate STEMs the tide

Every year we need a new flu vaccine, because influenza virus constantly mutates the major target of antibodies to flu: the "head" region of the viral hemagglutinin (HA) protein. Avoiding the problem of mutation requires a vaccine that elicits antibodies against the more conserved "stem" region of HA. During infection, antibodies are occasionally produced that recognize the stem and that neutralize a broad range of influenza virus strains. Impagliazzo *et al*. engineered an HA stem–only vaccine candidate that elicited broadly neutralizing antibodies in mice and nonhuman primates and that protected mice against multiple influenza strains. — KLM

Science, this issue p. 1301

APPLIED OPTICS

Wrap-around invisibility cloak

An invisibility cloak can be used to conceal an object from view by guiding light around it. Most cloaks developed so far have bulky structures that are difficult to scale up for hiding large objects. To design a thin invisibility cloak that can be wrapped around an object such as a sheet or skin, Ni *et al*. designed a two-dimensional metamaterial surface. Such flexible, highly reflective materials could be manufactured at large scale to hide large objects. — ISO

Science, this issue p. 1310

CANCER AND DEVELOPMENT

Smoothened signals through G proteins

During development, the Hedgehog signaling pathway regulates gene expression through the Gli family of transcription factors. The gene encoding the Hedgehog effector Smoothened (Smo) is aberrantly expressed in some types of breast cancers, and thus various drugs have been developed to target this protein. Villanueva *et al*. showed that Smo promoted cell proliferation in the mammary glands of mice via the G protein $G\alpha i_2$, rather than through Gli-mediated changes in gene expression (see also the Focus by Ogden). This discovery indicates that Smo-targeting drugs should also be screened for their ability to inhibit $G\alpha i_2$ activation. — WW

Sci. Signal. **8**, ra92 and fs16 (2015)

FERROELECTRICS

Thinning films induces ferroelectricity

Thin ferroelectric films are needed in computers and medical devices. However, traditional ferroelectric films typically become less and less polarized the thinner the films become. Instead of using a good ferroelectric and making it thinner, Lee *et al*. started with $SrTiO_3$, which in its bulk form is not ferroelectric. This material does have naturally occurring nanosized polarized regions. and when the thickness of the $SrTiO_3$ films reaches the typical size of these regions, the whole film aligns and becomes ferroelectric. — JS

Science, this issue p. 1314

ORGANIC CHEMISTRY

Lighting the way to aryl C-N bonding

Medicinal chemists like to add N bonds to the C atoms of aromatic rings to make bioactive compounds. By harnessing the energy in visible light, Romero *et al*. made these links and transformed C-H into C-N bonds. They used a blue-absorbing acridinium ion to activate a ring C for an incoming N partner. A nitroxyl radical co-catalyst (TEMPO) then choreographed the transfer of the H atom to O. The reaction worked for a broad range of substrates, including ammonium as a N source. — JSY

Science, this issue p. 1326

NEURODEVELOPMENT

Youthful damage limitation in stem cells

Every day brings more risk of damage to stem cells, which could have consequences for the whole organism. Moore *et al*. observed that dividing neural stem cells in rodents establish a diffusion barrier that restricts damaged proteins to one daughter cell, leaving the other with intact molecules. But with age this diffusion barrier weakens, so that replicating stem cells of older animals are less able to exclude damaged proteins than are the stem cells of younger rodents. — PJH

Science, this issue p. 1334

PROSTATE CANCER

Circulating signals of drug resistance

Cancer drugs often lose their effectiveness because tumors acquire genetic changes that confer drug resistance. Ideally, patients would be switched to a different drug before tumor growth resumes, but this requires early knowledge of how resistance arose. Miyamoto *et al*. have developed a non-invasive method to spot resistance by sequencing RNA transcripts in single circulating tumor cells (CTCs) (see the Perspective by Nanus and Giannakakou). For example, in prostate cancer patients, drug resistance was triggered by activation of the Wnt signaling pathway. But CTCs are rare and fragile, and the technology needs further development before it is used in clinical practice. — PAK

Science, this issue p. 1351; see also p. 1283

VIROLOGY

A good model for the "good boy virus"

Infection with human pegivirus (HPgV) provides a measure of protection to HIV-infected individuals by slowing HIV replication. This phenomenon has earned HPgV the nickname "good boy virus." How HPgV achieves this protective effect remains a mystery in part because no animal model of HPgV infection exists. Bailey *et al.* discovered that viruses closely related to HPgV also occur in wild baboons and can infect laboratory macaques, thus providing an opportunity to discover what tissues HPgV infects and how it is transmitted and replicated. — OMS

Sci. Transl. Med. **7**, 305ra144 (2015).

IN OTHER JOURNALS

Edited by **Sacha Vignieri** *and* **Jesse Smith**

Larval silkworms need an extra boost on their way to metamorphosing into adult moths

DEVELOPMENT

The path to metamorphosis

The metamorphosis of a caterpillar into a moth was classically thought to be controlled largely by two hormones: one that promotes metamorphosis and one that maintains juvenile characteristics. Daimon *et al*. definitively tested the role of the latter, juvenile hormone, in larval silk moths. They made knockout mutants that lacked an enzyme that catalyzes juvenile hormone synthesis or lacked the two juvenile hormone receptors. Analysis of these animals showed that juvenile hormone functioned only in the late larval stages, when it restrained metamorphosis until larvae reached a sufficient size. Competence to metamorphose depended not on release from juvenile hormone inhibition but rather on the accumulation of a yet-to-be-identified signal that controlled gene expression. — LBR

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1506645112 (2015).

NEUROSCIENCE

Cost and outcome shape habit

When rodents or monkeys perform a learned task, neurons in the striatum are active right before and after the performance. Their activity is thought to encode expected and experienced cost and outcome. These signals may be disrupted in Parkinson's disease, manifesting as problems with the initiation of movements. Desrochers *et al*. asked whether these neurons participate in the formation of habits in naïve monkeys. They therefore measured activity in the striatum while monkeys learned to solve a visual task without instructions

PHOTO: DAVID M. PHILLIPS/SCIENCE SOURCE

ECOLOGY

A jellyfish smörgåsbord

Jellyfish blooms are a common feature in oceans around the world. Because jellyfish hunt for zooplankton, crustaceans, and other lower-trophic-level prey, such blooms might disrupt food chains and have a negative impact on other marine predators. Sato *et al*. attached video and data loggers to thick-billed murres (*Uria lomvia*), diving seabirds that hunt in the Bering Sea. The seabirds successfully adapted their diving and hunting to feed on fish shoaling among the tentacles of large jellyfish, which suggests that blooms could have a beneficial impact on some seabirds. — GR

Biol. Lett. 10.1098/rsbl.2015.0358 (2015).

Thick-billed murres modify their hunting strategies when diving for fish among jellyfish blooms

("self-learning"). They found that before-and-after signals existed from the start and that a combination of cost and outcome signals afterward increased when monkeys got into the habit, and their performance improved. — LNS

Neuron 87, 853, 10.1016/j.neuron.2015.07.019 (2015).

HUMAN EVOLUTION

Old and new immune genetic variation

The Khloe-San peoples of Africa are one of the most ancient lineages of humans and thus can be examined to identify both ancient and recent human-specific genetic variation. Investigating the alleles of the killer cell immunoglobulin-like receptors and their human leukocyte antigen class I ligands, which are involved in the immune response, Hilton *et al*. found evidence of both old and new genetic variants. Because these genes can affect pregnancy, the identification of one recently evolved variant at relatively high frequencies suggests that it may have conferred a selective advantage. — LMZ

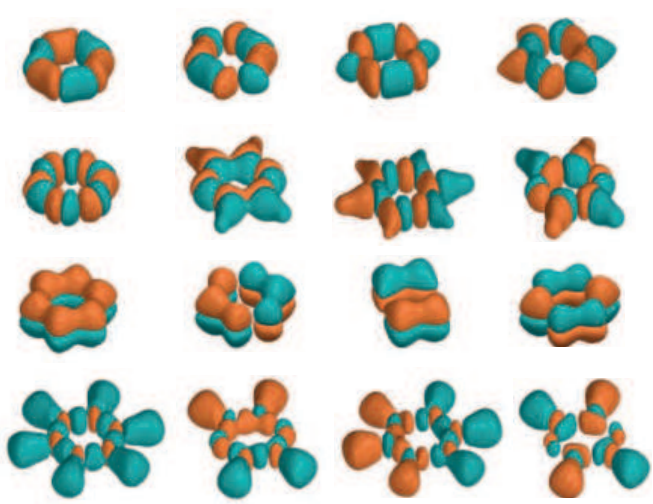
PLOS Genet. 10.1371/journal.pgen.1005439 (2015).

MOLECULAR PHYSICS

Molecular simulation

Determining the electronic structure and dynamics of molecular systems presents an intractable computational challenge for all but the simplest of molecules. However, the ability to tune the interactions between atoms trapped in an optical lattice provides a versatile quantum system with which to simulate complex condensed-matter systems. Luhmann *et al*. show that such a cold-atom system can be used to simulate the electronic dynamics of complex molecules and to generate three-dimensional high-resolution images of their molecular orbitals, using benzene as an example. — ISO

Phys. Rev. X **5**, 31016 (2015).



Calculated molecular single-particle orbitals of an artificial benzene molecule

MECHANOCHEMISTRY

Pulling for ketenes and imines

Expanding the range of reactive functional groups that can be generated by polymer mechanochemistry—literally pulling the chains apart by applying force—is important for creating new self-healing materials. Robb and Moore show that the Staudinger cycloaddition [the formation of a four-membered β-lactam ring from an imine (=N–R, where R is an alkyl or aryl group) and a ketene (=C=O)] could be reversed mechanically. Poly(methyl acrylate) polymers with an internal β-lactam ring were cleaved by ultrasonication in organic solvent. The highly reactive ketene was scavenged with isobutanol. — PDS

J. Am. Chem. Soc. 10.1021/jacs.5b07345 (2015).

ENVIRONMENTAL SCIENCE

Fertilizing water contamination

Excess nutrients such as nitrogen and phosphorus can negatively affect aquatic ecosystems. Nutrient-rich fertilizer runoff stimulates productivity, which can lead to harmful algal blooms or fish kills. Nolan and Weber show that such pollution has another wide-ranging consequence: increasing the mobility of uranium in groundwater. Geochemical data for two major aquifers in the United States, which in combination provide drinking water to millions and irrigation for one-sixth of U.S. agriculture, show that increased nitrate levels correlate strongly with the presence of uranium. Nitrate can oxidatively dissolve naturally occurring uranium minerals, which in turn can lead to potentially harmful levels of soluble uranium in groundwater, especially in shallow aquifers. — NW

Environ. Sci. Technol. Lett. 10.1021/acs.estlett.5b00174 (2015).

ALSO IN *SCIENCE* JOURNALS

Edited by **Caroline Ash**

ECONOMICS

Few thoughts for those with the most

A weighty scholarly tome has sparked a year-long public discussion of the unevenness of income and wealth distributions in the United States. In essence, a few people have a lot of both. Moral philosophers and economists have argued for centuries about the tradeoffs in life strategy that might explain wealth imbalance: between fairness and selfishness, and equality and efficiency. Fisman *et al.* describe the preferences of a group of elite students at Yale Law School. These elites lean toward selfishness and efficiency more than the average American, and these preferences are reflected in their job choices. — GJC

Science, this issue p. 1300

METALLIC GLASSES

Percolating cluster, fractal structure

Metallic glasses are appealing materials because they are strong and can bend without breaking. These materials are disordered but possess none of the defects seen in crystalline counterparts. Chen *et al.* developed a model for metallic glasses in which clusters of atoms float free in the liquid, begin to jam, and finally organize into a short-range fractal structure below the glass transition temperature. This model also accounted for the density and high strength characteristics of bulk samples. — BG

Science, this issue p. 1306

QUANTUM OPTICS

Entangling atoms by persistent poking

In quantum mechanics, repeated measurements targeting a particular unoccupied state of the system can keep that state from being occupied. Barontini *et al.* used this so-called quantum Zeno effect to restrict the dynamics of an ensemble of 36 ^{87}Rb atoms acting as qubits and residing in an optical cavity. The measurement of the cavity transmission blocked off the collective state in which all qubits were in their ground state. The ensuing dynamics resulted in the entanglement of the atoms, creating a potential resource for quantum information processing. — JS

Science, this issue p. 1317

CERAMIC FUEL CELLS

Cooler ceramic fuel cells

Ceramic ion conductors can be used as electrolytes in fuel cells using natural gas. One drawback of such solid-oxide fuel cells that conduct oxygen ions is their high operating temperatures (at least 600°C). Duan *et al.* have made a proton-conducting ceramic fuel cell with a modified cathode material that exhibits high performance on methane fuel at 500°C (see the Perspective by Gorte). — PDS

Science, this issue p. 1321; see also p. 1290

MINERAL SURFACES

Driving dissolution with x-rays

Carbonate minerals are important for Earth's carbon cycle. They precipitate directly from solution into diverse materials, depending on their physical or biological source. Whether carbonate minerals grow or dissolve is controlled by the thermodynamic drivers of the mineral/water interface. To control and observe the reactions, Lanaait *et al.* developed a synchrotron x-ray technique that images calcium carbonate surfaces in water and selectively tunes the solution saturation state (see the Perspective by Wolthers). The x-ray beam drives fast-moving reaction fronts far from equilibrium that are more limited by solution-ion transport than by surface processes. — NW

Science, this issue p. 1330; see also p. 1288

VIROLOGY

The devil in the dengue details

Along with their mosquito vectors, dengue viruses are spreading worldwide to infect millions of people. For a few, subsequent infection results in lethal hemorrhagic disease. Katzelnik *et al.* used antibody-binding data to map structural divergence and antigenic variation among dengue viruses. Comparing results in monkeys and humans, the viruses approximately clustered into the four known groups. However, the four virus groups showed as much antigenic distance within a group as between groups. This finding helps explain why immune responses to dengue are highly variable, and it has complex implications for epidemiology, disease, and vaccine deployment. — CA

Science, this issue p. 1338

STRUCTURAL BIOLOGY

Caught in the act of copying

The genomes of double-stranded RNA (dsRNA) viruses consist of about a dozen dsRNA segments enclosed by a protein coat. Inside the host cell, the coat remains intact, and the dsRNAs have to replicate within the coat. Liu and Cheng used cryo–electron microscopy of cypovirus particles to catch the dsRNAs in the act of being copied. The structures revealed that the RNA formed a liquid-crystalline array on which viral enzymes carry out multiple rounds of transcription to replicate the viral genome. — GR

Science, this issue p. 1347

SMALL PEPTIDES

Small peptide regulates protein activity

Coding and noncoding RNAs can produce peptides from small open reading frames (smORFs), with a variety of mostly unknown functions. Using a genome-wide screen, Zanet *et al.* show that Polished rice (Pri) smORF peptides control fruit fly development by binding to an E3 ubiquitin ligase. This changes the ligase's selectivity and triggers proteasome-dependent maturation of the developmental transcription factor Shavenbaby. Other smORF peptides may act by a similar mechanism to regulate protein activity. — BAP

Science, this issue p. 1356

ENERGY

A call for fossil fuel price reform

Coal is responsible for more carbon dioxide emissions per ton burnt than any other fossil fuel. Yet coal plants are being built around the world, particularly in China and other rapidly developing countries. In his Perspective, Edenhofer explores the reasons for the renaissance of coal. Fossil fuel subsidies are a key factor driving increased coal use, because they not only factor in fiscal spending but also include social costs, such as those to public health. Unless prices are reformed quickly to remove such wide-ranging incentives, climate change mitigation is imperiled. — JFU

Science, this issue p. 1286

RESEARCH ARTICLE SUMMARY

ECONOMICS

The distributional preferences of an elite

Raymond Fisman,* Pamela Jakiela, Shachar Kariv, Daniel Markovits

INTRODUCTION: Distributional preferences shape individual opinions and public policy concerning economic inequality and redistribution. We measured the distributional preferences of an elite cadre of Juris Doctor (J.D.) students at Yale Law School (YLS), a group that holds particular interest because they are likely to assume future positions of power and influence in American society. We compared the preferences of this highly elite group of students to those of a sample drawn from the American Life Panel (ALP), a broad cross-section of Americans, and to the preferences of an intermediate elite drawn from the student body at the University of California, Berkeley (UCB).

RATIONALE: We conducted modified dictator game experiments that varied the price of redistribution, i.e., the amount by which the "self's" payoff must be decreased in order to increase the payoff of the "other" (an anonymous other subject) by one dollar. In contrast to standard dictator games that do not vary the relative price of redistribution, our experimental design allows us to test whether our subjects' preferences are formally rational and to decompose subjects' preferences into two distinct tradeoffs: the tradeoff between self and other (fair-mindedness versus self-interest) and the tradeoff between equality and efficiency. For each subject, we estimated a constant elasticity of substitution (CES) utility function over payoffs to self and other; this functional form allows us to capture each tradeoff with a distinct parameter. A fair-minded subject places equal weight on the payoffs to self and other, whereas a selfish subject does not place any weight on the payoff to other; subjects' preferences may also fall in between these two extremes. A subject with distributional preferences weighted toward equality (reducing differences in payoffs) increases the expenditure share spent on other as the price of redistribution increases, whereas a subject with distributional preferences weighted toward efficiency (increasing total payoffs) decreases the expenditure share spent on other as the price of redistribution increases. An important strength of our measure of equality-efficiency tradeoffs between self and other is that it has been shown to predict such tradeoffs in distributional settings involving multiple others and to predict the likelihood of voting for political candidates perceived as favoring greater government redistribution. This work therefore captures, in an experimental setting, a plausible measure of subjects' attitudes toward actual redistributive policies.

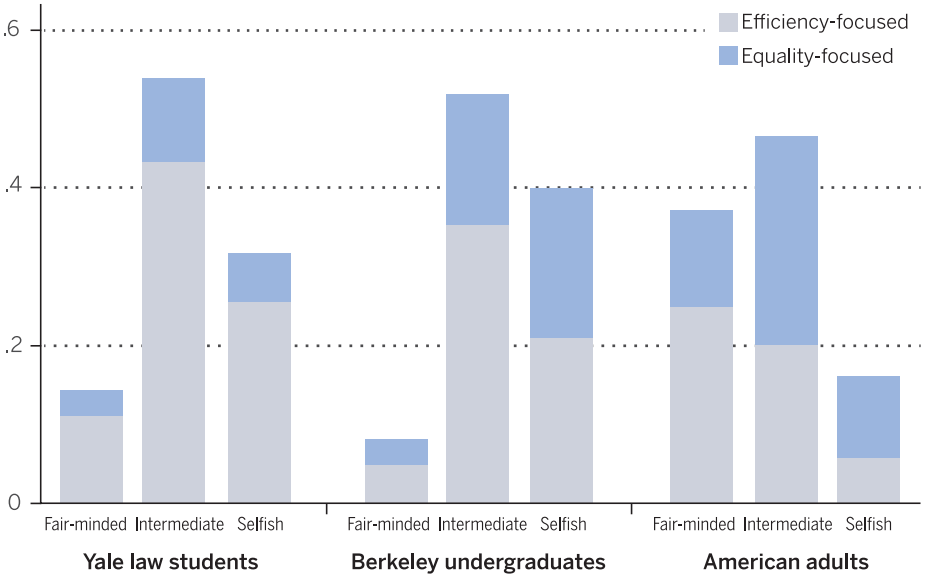
ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.aab0096

RESULTS: YLS subjects were substantially more efficiency-focused than were the ALP subjects drawn from the general population. Overall, 79.8% of YLS subjects were efficiency-focused, versus only 49.8% of the ALP sample. The YLS subjects displayed this distinctive preference for efficiency over equality in spite of overwhelmingly (by more than 10 to 1) self-identifying as Democrats rather than Republicans. In addition, YLS subjects were less likely to be classified as fair-minded and more likely to be classified as selfish than were the ALP subjects. Subjects from the intermediate elite fell between the YLS and ALP subjects with respect to efficiency-mindedness but were less likely to be fair-minded and more likely to be selfish than were the YLS subjects. We also demonstrate the predictive validity of our experimental measure of equality-efficiency tradeoffs by showing that it predicts the subsequent career choices of YLS subjects: More efficiency-focused behavior in the laboratory was associated with a greater likelihood of choosing private sector employment after graduation, whereas more equality-focused behavior was associated with a greater likelihood of choosing nonprofit sector employment.

CONCLUSION: Our findings indicate sharp differences in distributional preferences between subjects of varying degrees of eliteness. These results provide a starting point for future research on the distinct preferences of the elite and differences in distributional preferences across groups more generally. From a policy perspective, our results suggest a new explanation for the modesty of the policy response to the rise in income inequality in the United States: Regardless of party, the policymaking elite is significantly more focused on efficiency vis-a-vis equality than is the U.S. public. ■

Classifying subjects' distributional preferences. We classify subjects as either fair-minded, intermediate, or selfish and as either equality-focused or efficiency-focused. The bars show the fraction of subjects in each category of self-interest in the elite YLS, UCB (the intermediate elite), and relatively less elite ALP samples. Each bar is then split into equality-focused and efficiency-focused subgroups, denoted by blue and gray, respectively.



The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: rfisman@bu.edu
Cite this article as R. Fishman *et al.*, *Science* **349**, aab0096 (2015). DOI: 10.1126/science.aab0096

RESEARCH ARTICLE

ECONOMICS

The distributional preferences of an elite

Raymond Fisman,[1]* Pamela Jakiela,[2] Shachar Kariv,[3] Daniel Markovits[4]

We studied the distributional preferences of an elite cadre of Yale Law School students, a group that will assume positions of power in U.S. society. Our experimental design allows us to test whether redistributive decisions are consistent with utility maximization and to decompose underlying preferences into two qualitatively different tradeoffs: fair-mindedness versus self-interest, and equality versus efficiency. Yale Law School subjects are more consistent than subjects drawn from the American Life Panel, a diverse sample of Americans. Relative to the American Life Panel, Yale Law School subjects are also less fair-minded and substantially more efficiency-focused. We further show that our measure of equality-efficiency tradeoffs predicts Yale Law School students' career choices: Equality-minded subjects are more likely to be employed at nonprofit organizations.

Growing economic inequality has intensified interest in the distinctive attitudes and behaviors of the American elite, whose sense of entitlement increasingly captures both general and scholarly attention (*1*). The interest in elites is not just voyeuristic, but practical; elites, and in particular graduates of elite universities and professional schools, exert considerable influence over public and private policy in the United States today. For example, over the past century more than half of the presidents, including the past four, attended Yale, Harvard, or Princeton. The preferences of a relatively small number of current students will therefore have a large and highly disproportionate impact on the future of the country as a whole.

We studied the distributional preferences of an important pool of future elite policy-makers and citizens: Juris Doctor (J.D.) students at the Yale Law School (YLS). As Alexis de Tocqueville observed in the 19th century, lawyers constitute an American aristocracy and wield an outsized influence over society in general and public policy in particular. Tocqueville's observations remain true today, particularly as they pertain to a very small number of top law schools, of which YLS is the most selective. Although the American Bar Association does not rank law schools, YLS has been ranked first in the country by U.S. News and World Report every year since 1987, when it began publishing the ranking (*2*). We compared the distributional preferences of this elite group of students to those of a sample drawn from the American Life Panel (ALP), an internet survey of a diverse population of U.S. adults.

[1]Department of Economics, Boston University, Boston, MA, USA. [2]Department of Agricultural and Resource Economics, University of Maryland, College Park, MD, USA. [3]Department of Economics, University of California, Berkeley, Berkely, CA, USA. [4]Yale Law School, Yale University, New Haven, CT, USA.
*Corresponding author. E-mail: rfisman@bu.edu

Distributional preferences are important inputs into any measure of social welfare and enter every realm of policy-making. These preferences may naturally be divided into two qualitatively different components: the tradeoff between fair-mindedness and self-interest and the tradeoff between equality and efficiency. Although these two components of distributional preferences often operate together, they are conceptually distinct.

First, policy-makers must constantly decide whose interests matter and how much they matter. A baseline commitment to fair-mindedness—the ideal that all persons' interests matter equally—should inform all legitimate public policy. But in practice, fair-mindedness can be difficult to sustain against the many temptations to prefer one's own interests over the interests of others.

Second, policy-makers trade off equality and efficiency because reducing economic inequality almost inevitably has a cost; to use Okun's (*3*) famous metaphor, the transfer mechanisms that promote equality all involve leaky buckets. Policy-makers must thus decide, both in general and in any number of particular cases, by how much they are prepared to reduce aggregate income in order to secure a more equal income distribution. A comparison of the familiar philosophical theories of distributive justice—utilitarianism, for example, and Rawlsianism—further emphasizes the reasonable disagreements that fair-minded (impartial) policy-makers may have in trading off equality and efficiency.

In order to study the distributional preferences of an elite, we conducted laboratory experiments with YLS students using modified dictator games that vary the relative price of redistribution, building on the experiment first used by Andreoni and Miller (*4*). These decision problems are presented by using a graphical experimental interface that allows for the collection of rich individual-level data sets, as in (*5*). Specifically, we study a dictator game in which a subject divides an endowment between "self" and an anonymous "other." We denote persons self and other by s and o, respectively, and the associated monetary payoffs by π_s and π_o. In each decision problem, self allocates a unit endowment to π_s and π_o at fixed price levels p_s and p_o so that $p_s\pi_s + p_o\pi_o = 1$. This configuration creates budget sets over π_s and π_o in which $p = p_o/p_s$ is the relative price of redistribution.

The choice from a budget set indicates a subject's preferred allocation relative to a broad range of possible alternatives; it therefore provides more information about preferences than a choice from a discrete set of options would reveal. Furthermore, because of the user-friendly experimental interface, it is possible to present each subject with many choices in the course of a single experimental session, yielding an extremely rich data set. These data allow us to apply powerful techniques from demand analysis to determine whether each subject's behavior is consistent with utility maximization and to identify the structure of the utility function that rationalizes each subject's choices.

Our analysis examines the differences between the distributional preferences of elite YLS subjects—particularly, their willingness to sacrifice efficiency to reduce inequality—and the distributional preferences of the diverse sample of (relatively less elite) Americans in the ALP subject pool. In contrast to standard dictator games that do not vary the relative price of redistribution, our design allows us to separate fair-mindedness from equality-efficiency tradeoffs by examining subjects' responses to changes in the relative price of redistribution. A subject who decreases the expenditure share spent on other, $p_o\pi_o$, when the relative price of redistribution p increases has distributional preferences weighted toward efficiency (increasing total payoffs), whereas a subject who increases $p_o\pi_o$ when p increases has distributional preferences weighted toward equality (reducing differences in payoffs). For each subject, we constructed a measure of equality-efficiency tradeoffs by estimating a constant elasticity of substitution (CES) utility function over payoffs to self and other. A strength of this measure is that it has been shown to predict the equality-efficiency tradeoffs in distributional settings involving multiple others (*5*) and to predict the likelihood of voting for political candidates perceived as favoring greater government redistribution (*6*). We further validated the external validity of our measure in the present study by showing that it predicts YLS subjects' subsequent career choices. Taken together, this suggests that our measure of equality-efficiency tradeoffs meaningfully captures individual distributional preferences that govern subjects' real-world decisions.

Subject pools

YLS subjects

We conducted experimental sessions at YLS during the spring semesters of 2007, 2010, and 2013. The 3-year lag between experiments means that each set of sessions draws from an entirely

new YLS student body. Of the 208 subjects in the YLS sample, 199 reported their year of study; 91 subjects were 1st-year students, 61 were 2nd-year students, and the remainder were 3rd-year students. Summary statistics on the basic sociodemographic characteristics of our two main pools of subjects, the YLS and ALP samples, are reported in Table 1 (*7*).

YLS enrolls about 200 students per year, making it among the smallest and most selective graduate law schools in the United States. In the most recent year for which data are available, YLS accepted only 11.3% of its (already elite) college-educated applicants. YLS students tend to come from educated, relatively well-off households. In our experiments, 95 YLS subjects reported that both parents hold graduate degrees, and 113 grew up in U.S. ZIP codes where the average household income was above $70,000 in 2014 inflation-adjusted dollars (the mean household income in the U.S. was $72,641 in 2014). YLS students also have extremely high expected future incomes; although YLS does not disclose the starting salaries of its graduates, the median starting salary for graduates at top law schools such as Yale, Harvard, and Columbia is $160,000 per year (often augmented by signing bonuses). Overall, the YLS subjects are one of the most academically elite groups in the United States and can, in expectation, expect to join the ranks of the economic and political elite as well.

ALP subjects

For comparative purposes, we present our YLS data alongside a subset of the data of (*6*), collected in 2013 by using near-identical experiments with the ALP, an internet survey of more than 5000 adult Americans. The overall sample of ALP respondents is broadly comparable with the U.S. population in terms of demographic and socioeconomic characteristics; it includes an enormous amount of demographic, socioeconomic, and geographic diversity. Fisman *et al.* provide a detailed comparison of ALP subjects with respondents from the American Community Survey (ACS) conducted by the U.S. Census and representative of the U.S. population (*6*). The subsample of 1002 ALP respondents in the subject pool described in (*6*) is remarkably consistent with the entire ALP sample.

To focus on ALP subjects comparable in age with YLS subjects, we restricted attention to the 309 subjects in the original sample who were aged 40 and under. Summary statistics on the basic sociodemographics of the 309 subjects included in our analysis are reported in Table 1. As Table 1 indicates, the overwhelming majority of the ALP subjects are less educated than the YLS subjects.

Intermediate elites

We probed the generalizability of our results with the YLS and ALP samples by examining the behavior of two intermediate elites. This can help to rule out, in particular, the possibility that we are simply picking up a law school effect. First, we compared the most highly educated, wealthy ALP subjects to a nonelite comparison group of ALP subjects with less education and income. Second, we compared subjects drawn from the large and diverse student body of University of California, Berkeley (UCB) undergraduates to the full sample of ALP subjects (aged 40 and under). By examining these two intermediate elites, including one drawn from a broad cross-section of the general (primarily nonstudent) population, our aim was to assess the extent to which our conclusions are likely to reflect the distinctive distributional preferences of the U.S. elite, and not just those of YLS students or, more broadly, those in the legal profession.

ALP elite

We classify an ALP respondent as elite if she or he is employed, reported an annual household income over $100,000, and holds a graduate degree. In the experiments of Fisman *et al.*, only 9 of the 309 subjects aged 40 and under met these criteria (*6*). To obtain a larger sample of elite ALP subjects, we conducted an additional round of experiments in 2014, inviting all ALP respondents who met our criteria for eliteness to participate, along with a comparison group of nonelite ALP respondents (who were also employed and aged 40 and under but reported household incomes below $100,000 and did not hold graduate degrees). Combining the data on elite and nonelite ALP subjects from our two waves of experiments, we defined two additional samples: ALP elite (54 subjects) and ALP nonelite (206 subjects) (*8*).

UCB student elite

We also examined a second intermediate elite: undergraduate students at UCB, which is ranked as the world's top public university and among the most prestigious universities—public or private—globally. As a large public university, UCB draws its students from a diverse range of socioeconomic and cultural backgrounds. It is therefore a useful additional comparator for assessing both whether the patterns we attributed to the YLS subjects' eliteness hold more broadly, and also for emphasizing the extreme preferences we observed in the highly elite YLS sample. To this end, we used data collected in 2004 and 2011 in identical experiments at the UCB Experimental Social Science Laboratory (Xlab). The Xlab draws its subjects from all students and administrative staff, but most subjects in its experiments are undergraduate students. Fisman *et al.* describe the make-up of UCB student population during 2004–2011 and the composition of the UCB subjects in these experiments (*9*).

Table 1. Summary statistics on subjects in YLS and ALP samples.

	Subject pool	
	YLS subjects	ALP subjects
Age	25.40	31.23
Female	0.466	0.653
Born in the United States	0.782	0.906
Non-Hispanic white	0.632	0.545
Completed college	1	0.327
Observations	208	309

The experiment

In our experiments, we presented subjects with a sequence of 50 decision problems in which each choice has consequences for self (the subject) and for an anonymous other. Each decision problem is presented as a choice from a two-dimensional budget set. A choice of the allocation (x, y) from the budget set represents an allocation between accounts: Self received the tokens allocated to the y account, and other received the tokens allocated to the x account. More precisely, each decision involved choosing a point on a budget set over possible token allocations to self and other so that $p_s\pi_s + p_o\pi_o = 1$, where π_s and π_o correspond to the payoffs to self and other, respectively, and $p = p_o/p_s$ is the relative price of redistribution. In each decision problem, the computer selected a budget set at random from the set of budget sets that intersected at least one of the axes at 50 or more experimental tokens, but with neither intercept exceeding 100 tokens. These decision problems were presented by use of a graphical interface, and choices were made by using the mouse to move the pointer on the screen to the desired point (*10*). At the end of the experiment, one of each subject's choices was randomly selected to determine final payouts.

Framework for analysis

Nonparametric analysis

The most basic question to ask about choice data is whether it is consistent with individual utility maximization. If budget sets are linear (as in our experiment), classical revealed preference theory provides a direct test (*11*–*13*): Choices from a finite collection of budget sets are consistent with maximizing a piecewise linear, continuous, increasing, and concave utility function if and only if they satisfy the Generalized Axiom of Revealed Preference (GARP) (*14*). Hence, to assess whether an individual subject's choice data are consistent with utility-maximizing behavior, we needed to check whether the data satisfy GARP. Because our subjects make choices over a wide range of intersecting budget sets, our data provide a stringent test of utility maximization (*15*).

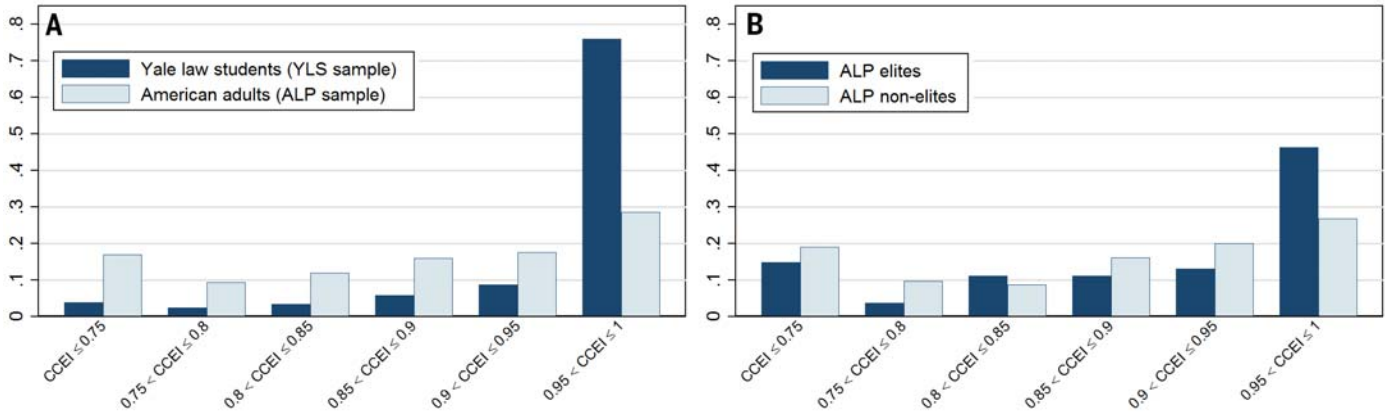


Fig. 1. CCEIs. (**A** and **B**) Histograms of the CCEI in (A) the YLS and ALP samples and (B) the ALP elite versus nonelite samples. CCEIs closer to 1 mean the data are closer to perfect consistency with GARP and hence to perfect consistency with utility maximization.

GARP provides a discrete test of utility maximization—either the data satisfy GARP or they do not—but individual choices frequently involve errors; subjects may compute incorrectly, execute intended choices incorrectly, or err in other less obvious ways. To account for the possibility of errors, we assessed how nearly individual choice behavior complies with GARP by using Afriat's (*16*) critical cost efficiency index (CCEI), which measures the fraction by which each budget constraint must be shifted in order to remove all violations of GARP. By definition, the CCEI is between 0 and 1: indices closer to 1 mean that the data are closer to perfect consistency with GARP and hence to consistency with utility maximization.

Parametric analysis

In the case of two goods, consistency with GARP and budget balancedness together imply that the demand function is homogeneous of degree zero. If we also assume separability and homotheticity, then the underlying utility function must be a member of the constant elasticity of substitution (CES) family (*17*). The CES utility function is given by

$$u_s(\pi_s, \pi_o) = [\alpha(\pi_s)^\rho + (1-\alpha)(\pi_o)^\rho]^{1/\rho}$$

where α represents the relative weight on the payoff for self vis-a-vis other (fair-mindedness), and ρ represents the curvature of the indifference curves (equality-efficiency tradeoffs).

Those with $\alpha = 1/2$ are fair-minded in the sense that they place equal weight on the payoffs to self and other; those with $\alpha = 1$ are perfectly selfish and do not put any weight on the payoff to other. Those with $1/2 < \alpha < 1$ exhibit some (intermediate) degree of fair-mindedness. For any $\rho > 0$, an increase in the relative price of redistribution raises—and for any $\rho < 0$, and increase in the relative price of redistribution lowers—the expenditure share on tokens allocated to self, $p_s\pi_s$. Thus, those with $\rho > 0$ have distributional preferences weighted toward maximizing efficiency (increasing total payoffs), whereas those with $\rho < 0$ have distributional preferences weighted toward minimizing inequality (reducing differences in payoffs).

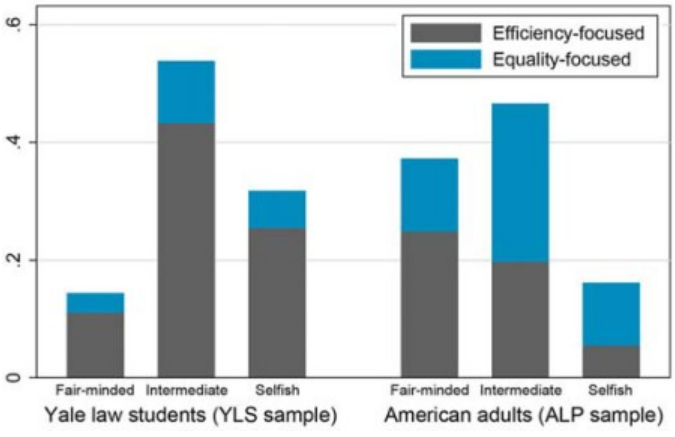


Fig. 2. Classifying subjects' distributional preferences. We classify subjects as either fair-minded ($\hat{\alpha}_n < 0.95$), or selfish ($\hat{\alpha}_n > 0.95$) and as either equality-focused ($\hat{\rho}_n < 0$) or efficiency-focused ($\hat{\rho}_n > 0$). The bars show the fraction of subjects in each category of self-interest in the YLS and ALP samples. Each bar is then split into equality-focused and efficiency-focused subgroups, denoted by blue and gray, respectively.

Our estimation was done for each subject n separately, generating individual-level estimates of the CES parameters. Specifically, we normalized prices at each observation and estimated demand in terms of budget shares, which are bounded between 0 and 1, using nonlinear Tobit maximum likelihood.

Experimental results

In this section, we provide results from the YLS and ALP samples. We first examine whether the data observed in our experiment could have been generated by a subject maximizing a well-behaved utility function. We then proceed to our econometric analysis by imposing further structure on the data in order to recover the underlying distributional preferences.

Rationality

The mean CCEI in the YLS sample is 0.95, and the median is 0.99, indicating that the overwhelming majority of the YLS subjects make choices that are perfectly or almost perfectly consistent with utility maximization. For comparison, the mean CCEI in the ALP sample is 0.86, and the median is 0.89. Thus, the choices of the ALP subjects are generally consistent with utility maximization (*18*). Nonetheless, the CCEIs of the YLS subjects are substantially higher than those of subjects in the ALP sample. A Wilcoxon rank-sum test rejects the hypothesis that the distributions of CCEIs are equal ($P < 0.001$). Histograms of CCEIs of both the YLS and ALP subjects are presented in Fig. 1A. Relative to the CCEIs of the general population in the ALP sample, the CCEIs of YLS subjects are skewed to the right (*19*).

Preferences

Our subjects' CCEIs are sufficiently near 1 to justify treating the data as utility-generated. If we also assume separability and homotheticity, then the underlying utility function $u_s(\pi_s,\pi_o)$ that rationalizes the data must be a member of the CES family. We now turn to the analysis of our estimates of the individual CES utility parameters, $\hat{\alpha}_n$ and $\hat{\rho}_n$. The distributions of $\hat{\alpha}_n$ and $\hat{\rho}_n$ in the YLS and the ALP samples are summarized in Fig. 2. Across all categories of self-interest (fair-minded, intermediate, and selfish), the YLS

subjects are substantially more efficiency-focused than are the ALP subjects drawn from the general population. Overall, 79.8% of YLS subjects are efficiency-focused ($\hat{\rho}_n > 0$) versus only 49.8% of the ALP sample. In addition, the YLS subjects are less likely to be classified as fair-minded and more likely to be classified as selfish than are the ALP subjects: 14.4% of YLS subjects are classified as fair-minded, as compared with 37.2% of ALP subjects; conversely, 31.7 and 16.2% of YLS and ALP subjects, respectively, are classified as selfish (*20*).

The distributions of the parameter estimates $\hat{\alpha}_n$ and $\hat{\rho}_n$ for the both YLS and ALP samples are presented in Figs. 3 and 4, respectively. The distribution of the estimated $\hat{\alpha}_n$ parameters of the YLS sample is skewed sharply to the right, relative to the distribution of the ALP sample (Fig. 3A); the YLS subjects are substantially less likely to have estimated $\hat{\alpha}_n$ parameters below 0.6 (22.1 versus 52.4%), are somewhat more likely to have estimated $\hat{\alpha}_n$ parameters between 0.6 and 0.9 (42.3 versus 27.5 %), and substantially more likely to have estimated $\hat{\alpha}_n$ parameters above 0.9 (35.6 versus 20.1 %). As shown in Fig. 4A, the distribution of the estimated $\hat{\rho}_n$ parameters of the YLS sample lies clearly to the right of the ALP sample's distribution of $\hat{\rho}_n$ values, indicating a much higher degree of efficiency orientation in our elite sample. A substantial majority of YLS subjects have estimated $\hat{\rho}_n$ parameters above 0.5 (60.1% of subjects), indicating a very high degree of efficiency focus; for comparison, only 17.8% of ALP subjects have estimated $\hat{\rho}_n$ parameters that high.

We next turned to regression analyses that more systematically examine the differences in fair-mindedness ($\hat{\alpha}_n$) and equality-efficiency tradeoffs ($\hat{\rho}_n$) between the YLS and ALP samples. We defined an indicator variable to denote the YLS sample and present the results of individual-level regressions with this as the primary explanatory variable in Table 2, which includes the results with no individual-level controls and when we control for gender, age, and education level (having a college degree).

In the first column of Table 2, we present a specification with the fair-mindedness parameter as the dependent variable, using a Tobit model that allows for the censoring of $\hat{\alpha}_n$ at 1. The parameters are, on average, 0.12 higher in the YLS sample than in the ALP sample ($P < 0.001$), indicating that the YLS subjects are substantially and significantly less fair-minded than the ALP subjects. After controlling for age, gender, and education level, we still find that YLS subjects are significantly less fair-minded than are ALP subjects ($P = 0.002$), but the point estimate drops to 0.084.

We next present a specification with the equality-efficiency tradeoff parameter as the dependent variable. Because the distribution of $\hat{\rho}_n$ is highly skewed ($\hat{\rho}_n$ ranges from $-\infty$ to 1 and a number of subjects have very low values of $\hat{\rho}_n$), we estimate quantile regressions that are less sensitive to extreme values. We report results for the 25th, 50th, and 75th percentiles in second through fourth columns of Table 2, respectively. Across all three quantiles, the values of $\hat{\rho}_n$ are significantly higher in the YLS sample than in the ALP sample ($P < 0.001$), indicating that the YLS subjects are significantly more focused on efficiency vis-a-vis equality than are the ALP subjects. After adding controls, the point estimates on YLS are reduced by about one half, but for both the 50th and 75th percentiles, the coefficient remains significant ($P = 0.003$ and 0.001, respectively).

As an alternative approach to dealing with the skewed distribution of $\hat{\rho}_n$, in the fifth column of Table 2 we present a probit specification using an

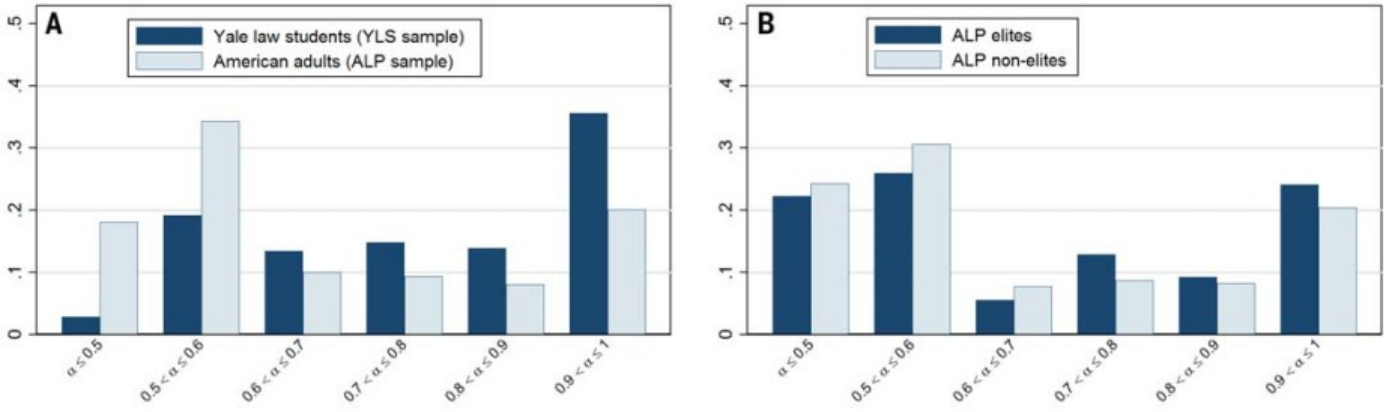


Fig. 3. Estimated $\hat{\alpha}_n$ parameters. (**A** and **B**) Histograms of the $\hat{\alpha}_n$estimates in (A) the YLS and ALP samples and (B) the ALP elite versus nonelite samples. $\hat{\alpha}_n$ indexes fair-mindedness: the relative utility weight placed on one's own payoff vis-à-vis the payoff to other.

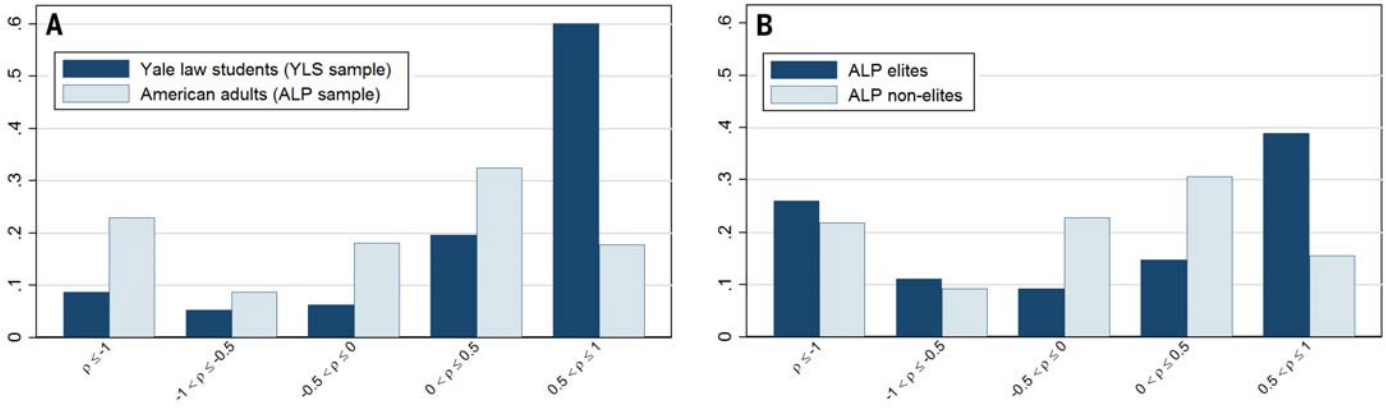


Fig. 4. Estimated $\hat{\rho}_n$ parameters. (**A** and **B**) Histograms of the $\hat{\rho}_n$ estimates in (A) the YLS and ALP samples and (B) the ALP elite versus nonelite samples. $\hat{\rho}_n$ indexes equality-efficiency tradeoffs; $\hat{\rho}_n$ values closer to 1 indicate greater efficiency focus.

Table 2. Regressions of estimated CES parameters, by subject pool. Standard errors are in parentheses. ***, significance at the 99% level; **, significance at the 95% level; *, significance at the 90% level.

		Quantile regressions			
Specification	**Tobit**	**25th percentile**	**50th percentile**	**75th percentile**	**Probit**
Dependent variable	$\hat{\alpha}_n$	$\hat{\rho}_n$	$\hat{\rho}_n$	$\hat{\rho}_n$	$I(\hat{\rho}_n > 0)$
Without controls					
YLS student	0.120***	1.075***	0.693***	0.405***	0.831***
	(0.017)	(0.217)	(0.080)	(0.049)	(0.122)
Constant	0.668***	−0.867***	0.005	0.420***	0.004
	(0.011)	(0.137)	(0.051)	(0.031)	(0.071)
Observations	517	517	517	517	517
Including controls for age, gender, and education level					
YLS student	0.084***	0.552	0.357***	0.260***	0.440**
	(0.027)	(0.476)	(0.120)	(0.077)	(0.183)
Constant	0.660***	−0.051	0.260**	0.604***	0.497***
	(0.029)	(0.480)	(0.121)	(0.078)	(0.183)
Observations	514	514	514	514	514

indicator for efficiency-oriented subjects ($\hat{\rho}_n > 0$) as the dependent variable. We found that the YLS subjects are 29.2 percentage points more likely to be efficiency-focused than are the ALP subjects ($P < 0.001$). After controlling for demographics, the YLS subjects are still 14.1 percentage points more likely to be efficiency-focused than are the ALP subjects ($P = 0.016$).

Our results are thus robust to the inclusion of controls for age, gender, and education. Education is a defining feature of the elite, and as such, whether it should be accounted independently for its role is unclear. Still, we argue in the spirit of Altonji *et al.* (*21*) that if unobserved attributes—which we would have expected a priori to be correlated with education—were a dominant source of the observed correlation between elite status and distributional preferences, then adding controls should have had a substantial effect on the estimated associations.

External validity—YLS subjects' career choices

Our results above show that subjects drawn from the student population at YLS—the future U.S. elite—are much more rational (in the sense of implementing a consistent, complete, and transitive preference ordering) and are far more inclined to favor efficiency over equality relative to subjects drawn from the ALP, a diverse cross-section of Americans. Yet, this analysis rests on the assumption of external validity; we assume that our individual-level laboratory experimental measure of equality-efficiency tradeoffs predicts the willingness to trade off equality and efficiency outside the laboratory. As discussed above, our experimental design was selected in part because it has been shown to predict equality-efficiency tradeoffs in a range of experimental settings (*5*) and to predict voting behavior (*6*). To further assess the external validity of our experimental measure of equality-efficiency tradeoffs, we tested whether YLS subjects' distributional preferences, as captured in our experiment, are reflected in behavior in a natural decision environments by looking at subjects' (early) career choices.

In late 2014, we obtained approval to access the names of subjects in the first two waves of the YLS experiment fielded in 2007 and 2010 (subjects who participated in 2013 are still students at YLS or at extremely early stages of their careers). We were able to track down, via Web searches, the career choices of 137 out of the 139 subjects (*22*). Of these, 119 subjects (86.9%) could be cleanly classified based on employer type: nonprofit (33 subjects), academia (13 subjects), government (18 subjects), and corporate (66 subjects). Of the remaining 17 subjects, 14 subjects extended their training as clerks, a position that can serve as preparation for a range of legal careers, and three continued their schooling.

YLS graduates who chose nonprofit careers tended to pursue the equality-related rights and interests of the disenfranchised. In contrast, YLS graduates who work in the corporate sector overwhelmingly serve as managers or deal-makers whose basic purpose is to extract efficiencies on behalf of their employers or clients. Moreover, although this observation is perhaps more anecdotal, the corporate workplace itself is more single-mindedly structured around efficiency than are workplaces in the nonprofit sector. We grouped government and academia as an intermediate case and examined whether the nonprofit and corporate subsamples have substantially and significantly lower and higher, respectively, efficiency orientations relative to other YLS subjects.

We assert that the existence of a relationship between our experimental measure of the equality-efficiency tradeoffs of YLS subjects and their real-world career choices would confer substantial external validity on the conclusions drawn from our laboratory experiments (*23*). The median $\hat{\rho}_n$ parameter value among YLS subjects employed in the nonprofit, academia/government, and corporate sectors are 0.439, 0.648 and 0.745, respectively. A Wilcoxon rank-sum test rejects the hypothesis that the nonprofit and corporate subsamples have equal $\hat{\rho}_n$ distributions ($P = 0.057$). A rank-sum tests rejects equality of the $\hat{\rho}_n$ distributions in the corporate and academia/government subsamples ($P = 0.035$) but (unsurprisingly given the small sample sizes) does not reject the hypothesis that the $\hat{\rho}_n$ distributions in the nonprofit and academia/government subsamples are equal ($P = 0.637$).

Last, we further investigated the relationship between the equality-efficiency tradeoffs of YLS subjects and their career choices using an ordered logit regression, ranking the (ascending) efficiency orientation of employment types as nonprofit, academia/government, or corporate. We report the results in Table 3. Given the skewed distribution of the estimated $\hat{\rho}_n$ parameters, we provide two alternative specifications: in Table 3, column 1, the independent variable is an indicator for having an above median (within the YLS sample) $\hat{\rho}_n$ value, whereas in column 2 it is the decile of the estimated $\hat{\rho}_n$ distribution. Table 3 also shows the results both with no individual-level controls and when we control for gender, age at the time the subject participated in the experiment, and the year of participation (either 2007 or 2010). The estimation results confirm our findings above in a regression setting: Our experimental measure of equality-efficiency tradeoffs predicts YLS subjects subsequent career choices.

Intermediate elites

We last examined whether our findings on the distinct distributional preferences of YLS subjects can plausibly be applied to elites more broadly. This analysis helps to ensure that we are not simply picking up an effect peculiar to the YLS population. To do so, we examined the behaviors of two intermediate elites who participated in identical experiments. First, we compared the most highly educated, wealthy ALP

Table 3. Ordered logit estimation of YLS subjects' career choices. Standard errors in parentheses. ***, significance at the 99% level; **, significance at the 95% level; *, significance at the 90% level. Dependent variable is equal to 1 for subjects who work in the nonprofit sector, equal to 2 for subjects who work in academia or government, and equal to 3 for subjects who work in the corporate sector. Controls are for age, gender, and year of experimental session.

Dependent variable: post-YLS career category		
	Without controls	
Above median $\hat{\rho}_n$	1.043***	
	(0.364)	
Decile of estimated $\hat{\rho}_n$		0.157**
		(0.068)
Observations	120	120
	With controls	
Above median $\hat{\rho}_n$	1.035***	
	(0.374)	
Decile of estimated $\hat{\rho}_n$		0.164**
		(0.076)
Observations	118	118

respondents with a comparison group of nonelite ALP respondents with less education and income. Second, we compared UCB undergraduate students with the ALP respondents drawn from the general population. By examining two different intermediate elites, including one drawn from a broad cross-section of the general (primarily nonstudent) population, our aim was to assess the extent to which our conclusions are likely to reflect the distinctive distributional preferences of the U.S. elite, and not just those of YLS students.

Elite ALP subjects

We classify an ALP subject as elite if she or he is employed, has an annual household income of over $100,000, and holds a graduate degree (although this definition does not approach the eliteness of the YLS subjects). We compared the elite ALP subjects with nonelite ALP subjects who are employed but with incomes below $100,000 and no graduate degree. The ALP subsample we used for this elite versus nonelite analysis comprises data collected across two waves of experiments and includes 54 ALP elite and 206 ALP nonelite subjects. Paralleling our main analysis, histograms comparing CCEI scores, fair-mindedness ($\hat{\alpha}_n$), and equality-efficiency tradeoffs ($\hat{\rho}_n$) between the ALP elite and ALP nonelite subsamples are presented in Figs. 1B, 3B, and 4B, respectively.

As shown in Fig. 1B, the distribution of CCEI scores for the ALP elite subjects is skewed to the right relative to that of the ALP nonelites. The mean and median CCEI scores of ALP elites are 0.882 and 0.945 versus 0.857 and 0.891 for nonelites, and a Wilcoxon rank-sum test rejects the equality of the distributions ($P = 0.014$). A much more modest association between ALP elite status and fair-mindedness ($\hat{\alpha}_n$) is shown in Fig. 3B—the mean and median values are 0.687 and 0.612 versus 0.659 and 0.576 for elites and nonelites, respectively. This indicates, as with our main analysis, a lower level of fair-mindedness among the elite, although here the difference in distributions is modest and statistically insignificant (Wilcoxon rank-sum test $P = 0.429$).

Last, there are sharp differences between the ALP elite and nonelite subjects in their equality-efficiency tradeoffs ($\hat{\rho}_n$) (Fig. 4B). Although we cannot reject the equality of the distributions (Wilcoxon rank-sum test $P = 0.209$), a much larger fraction of ALP elite subjects have high values of $\hat{\rho}_n$, indicating greater efficiency focus. Specifically, 38.9% of the ALP elite subjects have estimated $\hat{\rho}_n$ parameters of at least 0.5, as compared with only 15.5% for the ALP nonelites. In contrast, only 24.1% of the ALP elite subjects have intermediate $\hat{\rho}_n$ parameters between −0.5 and 0.5 as compared with 53.4% of ALP nonelite subjects. Thus, as with the YLS elite, we observed a stronger efficiency orientation among elites than among nonelites within the ALP subject pool (*24*).

Elite UCB subjects

We next turned to a second intermediate elite: undergraduates at UCB. Again paralleling our main analysis, histograms comparing the CCEIs, fair-mindedness ($\hat{\alpha}_n$), and equality-efficiency tradeoffs ($\hat{\rho}_n$) between the UCB subjects and the ALP subjects drawn from the general population are shown in figs. S1, S2, and S3. Overall, the differences between the ALP subjects and the UCB subjects are very similar to the differences between the ALP and YLS subjects.

A sharp difference in CCEIs between the UCB subjects and ALP subjects is shown in fig. S1, although the difference is slightly smaller than that observed in comparing the CCEIs of the YLS subjects with those of the ALP subjects. A Wilcoxon rank-sum test rejects the equality of the distributions ($P < 0.001$). Turning to the distributions of fair-mindedness ($\hat{\alpha}_n$), there is an even larger gap between the UCB and ALP subjects than between the YLS and the ALP subjects (reported in our main analysis); the median values for the UCB and ALP subject pools are 0.888 and 0.591, respectively (Wilcoxon rank-sum test $P < 0.001$) (fig. S2).

Last, there is a sharp difference between the UCB and ALP subjects in their equality-efficiency tradeoffs ($\hat{\rho}_n$) (fig. S3). Although not as great as the difference in $\hat{\rho}_n$ values for the YLS and ALP subjects, the UCB subjects' $\hat{\rho}_n$ values are skewed to the right relative to ALP subjects, indicating a greater efficiency orientation. The median $\hat{\rho}_n$ values are 0.259 and 0.005 for UCB and ALP subjects, respectively; a Wilcoxon rank-sum test rejects equality of the distributions ($P < 0.001$) (*25*).

Concluding remarks

People from all walks of life implement their distributional preferences in the real world. This is especially true for the elite YLS students in our sample, many of whom will assume positions of substantial power in economic and political affairs. We decomposed distributional preferences into two qualitatively different components—fair-mindedness and equality-efficiency tradeoffs—and measured both at the individual level in diverse samples of varying degrees of eliteness. Our experiment enabled us to distinguish fair-mindedness from equality-efficiency tradeoffs in the laboratory and to assess the extent of efficiency orientation in subject pools with different degrees of eliteness.

The increase in wealth and income inequality within and across countries is one of the defining social, economic, and political challenges of our time. Our results offer a potential new explanation for the muted policy response to increased income inequality in the United States: The equality-efficiency tradeoffs of the policy-making elite are such that they are far less inclined than is the general population to sacrifice efficiency to promote equality. As Gilens and Page (*26*) found, the preferences of the economic elites are far more correlated with public policy choices than are the preferences of the general public. Although there are many factors that contribute to the limited distributional response to rising inequality in the United States (ranging from loss-aversion, to attitudes toward fair-treatment of oneself by others, to moral hazard concerns, to beliefs about the extent of inequality), we focus on one potential cause: By favoring efficiency over equality, policy-makers may be acting on their own distributional preferences, which may be closely aligned with the interests and preferences of other members of the elite. The connections we draw between laboratory results and the degrees of eliteness promise to help in understanding the policies and practices that are implemented by the elite or the establishment in the broader world, and the experimental techniques and results that we have already developed provide promising tools for future work in this area.

Our results contribute to the broader discussion of the interplay between distributional preferences and tax policy. The vast and growing body of work on this topic includes theoretical analyses, experiments in the laboratory

and in the field, and surveys of distributional preferences, as well as the related but distinct notion of preferences for government redistribution. A recent study emphasizes the critical role that distributional preferences should play as a determinant of distributive policies generally and optimal tax policy in particular (*27*). This entire body of work overlooks a critical mediating factor between the measured distributional preferences of the general population and implemented policy choices: namely, the extent to which the distributional preferences of those in power differ from the preferences of voters. Ours is the first study to emphasize the existence of such differences, laying the groundwork for better understanding why public policy outcomes may diverge from stated voter preferences. We thus open a new and important window onto familiar mysteries, including the modesty of the policy response to rising inequality even in the face of growing popular outrage (*28*).

REFERENCES AND NOTES

1. The term "affluenza" has entered the English language to describe the phenomenon of growing elite entitlement and excessive consumption, and academic work has connected higher socioeconomic class to a greater expressed sense of entitlement and greater narcissism (*29*). We conceptualize elite status in terms of educational attainment and lifetime wealth.
2. All nine sitting Supreme Court justices and two of the past three presidents (as well as a frontrunner to become the next president) are graduates of either Yale or Harvard Law Schools. Most YLS students are also destined for membership in the economic elite, creating another potential channel of policy influence.
3. A. M. Okun, *Equality and Efficiency: The Big Tradeoff* (Brookings Institution Press, Washington, DC, 1975).
4. J. Andreoni, J. Miller, Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* **70**, 737–753 (2002). doi: 10.1111/1468-0262.00302
5. R. Fisman, S. Kariv, D. Markovits, Individual preferences for giving. *Am. Econ. Rev.* **97**, 1858–1876 (2007). doi: 10.1257/aer.97.5.1858
6. R. Fisman, P. Jakiela, S. Kariv, The distributional preferences of Americans. *National Bureau of Economic Research* **20145** (2014); available at www.nber.org/papers/w20145.
7. Detailed demographic data are not available for the YLS student body as a whole, but the evidence that we do have indicates that our YLS sample is broadly representative of the student body. Of YLS students graduating between 2007 and 2015, 46.7% were women, and 46.6% of our YLS subjects are female. Between 2007 and 2015, the average age at graduation was 28, whereas the average age in our YLS sample is 25.4 (but subjects are drawn from all 3 years of law school, and 1st-year students are slightly overrepresented).
8. In our ALP elites versus nonelites analysis, we omit 157 of the 309 ALP subjects aged 40 and under from the experiment of (*6*) who did not report their household incomes, provided inconsistent responses to the two-part income question in the ALP survey, were unemployed, or did not meet our criteria for (either) eliteness or noneliteness (for example, because they held graduate degrees but reported low household incomes).
9. R. Fisman, P. Jakiela, S. Kariv, How did distributional preferences change during the great recession? *J. Public Econ.* **128**, 84–95 (2015). doi: 10.1016/j.jpubeco.2015.06.001
10. Materials and methods are available as supplementary materials on *Science* Online.
11. S. N. Afriat, The construction of utility functions from expenditure data. *Int. Econ. Rev.* **8**, 67–77 (1967). doi: 10.2307/2525382
12. H. R. Varian, The nonparametric approach to demand analysis. *Econometrica* **50**, 945–972 (1982). doi: 10.2307/1912771
13. H. R. Varian, Non-parametric tests of consumer behavior. *Rev. Econ. Stud.* **50**, 99–110 (1983). doi: 10.2307/2296957
14. Let $\{(p^i, \pi^i)\}_{i=1}^{50}$ be the data generated by some individual's choices, where p^i denotes the ith observation of the price vector, and π^i denotes the associated allocation. An allocation π^i is directly revealed preferred to π^j, denoted $\pi^i R^D \pi^j$, if $p^i \cdot \pi^i \geq p^i \cdot \pi^j$. An allocation π^i is revealed preferred to a π^j, denoted $\pi^i R \pi^j$, if there exists a sequence of allocations $\{\pi^k\}_{k=1}^{K}$ with $\pi^1 = \pi^i$ and $\pi^K = \pi^j$, so that $\pi^k R^D \pi^{k+1}$ for every $k = 1,\ldots K-1$. In this notation, GARP requires that if $\pi^i R \pi^j$, then $p^j \cdot \pi^j \leq p^j \cdot \pi^i$ (that is, if π^i is revealed preferred to π^j, then π^j must cost at least as much as π^i at the prices prevailing when π^j is chosen).
15. Choi *et al.* (*30*) provide more details on the power of the GARP test. Bronars (*31*) builds on Becker (*32*) to provide a test based on a comparison of the behavior of actual subjects to the behavior of simulated subjects who randomize uniformly on each budget line. The power of Bronars' test is defined to be the probability that a random subject violates GARP. In our experimental design, all random subjects had violations, implying that the Bronars criterion attains its maximum value.
16. S. N. Afriat, Effciency estimates of production functions. *Int. Econ. Rev.* **13**, 568–598 (1972). doi: 10.2307/2525845
17. The multiple decisions that subjects make could lead to their conceiving of their choices as a portfolio. As a result, equality-focused subjects might make choices that are less equality-focused by allocating more than their conception of fairness requires to self if the relative price of redistribution exceeds one and more to other if the relative price of redistribution is less than one, evaluating their portfolio of choices according to its expected payoff. Because a single decision round is used for payoffs, an equality-focused subject that follows this approach will create extreme inequality ex post. This approach also is not supported by the experimental data; very few subjects allocate all of the tokens to self when the relative price of redistribution is greater than one, or none of the tokens to self when the relative price of redistribution is less than one. More generally, it is possible that a subject implementing a preference over the rounds approached as a portfolio might appear, on account of our focusing on separate rounds, to be implementing a different preference. But only unusual preference orderings have the property that when implemented over portfolios, they appear as the consistent implementation of different preferences round by round. Many preferences over portfolios will instead generate inconsistent choices analyzed round by round. Our subjects' almost uniformly high CCEI scores thus constitute strong evidence against the hypothesis that they approach the rounds as a portfolio.
18. Following Bronars (*31*), we compared the behavior of our actual subjects with the behavior of simulated subjects who randomize uniformly on each budget line. The mean CCEI in a sample of 25,000 simulated subjects is only 0.60; the vast majority of YLS subjects and many ALP subjects have CCEIs above 0.90, whereas none of the simulated subjects had a CCEI that high.
19. We found similar results when we screened out selfish subjects who allocate themselves an average of more than 95% of the tokens.
20. We found nearly identical results when different $\hat{\alpha}_n$ thresholds were used to classify subjects as fair-minded and selfish. Because the CES form approaches the Cobb-Douglas function $\pi_s^{\alpha}\pi_o^{1-\alpha}$ as $\rho \to 0$ (so that the expenditure shares are constant for any price of redistribution p), we only consider $\hat{\alpha}_n = 0$ as the threshold between preferences weighted toward efficiency versus equality.
21. J. G. Altonji, T. E. Elder, C. R. Taber, Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Polit. Econ.* **113**, 151–184 (2005). doi: 10.1086/426036
22. Of the two with no career data, one had, according to LinkedIn, left his or her last place of employment in 2011 but provided no further information. For the second subject, we could find no information whatsoever.
23. We did not perform a parallel analysis for the parameter measuring fair-mindedness. Fair-mindedness may implicate payoffs beyond monetary income. Although elite jobs in the nonprofit, government, and academic sectors pay less than elite corporate-sector jobs pay, they may offer substantial compensating personal benefits in the form of reputation. We therefore do not think it sensible to associate corporate employment with selfishness and nonprofit employment with fair-mindedness.
24. In table S1, we replicate the regression analyses reported in Table 2 with the ALP elite versus nonelite samples. We did not find a statistically significant association between ALP elite and equality-efficiency tradeoffs at the 50th percentile ($P = 0.224$), although we found a statistically significant association at the 75th percentile ($P = 0.016$). The differences between elites and nonelites are also insignificant when we look at CCEI scores and estimated $\hat{\alpha}_n$ parameters, but the coefficients are directionally consistent with higher consistency and lower fair-mindedness among elite subjects. Adding controls for gender and age has little effect on the coefficient estimates or significance levels.
25. In table S2, we replicate the regression analyses reported in Table 2 using the UCB and ALP samples. The differences between the UCB and ALP subjects are always in the same direction as the differences between the YLS and ALP subjects and are always statistically significant. We cannot include controls because we were unable to obtain individual-level data on UCB subjects owing to privacy concerns.
26. M. Gilens, B. I. Page, Testing theories of American politics: Elites, interest groups, and average citizens. *Perspect. Polit.* **12**, 564–581 (2014). doi: 10.1017/S1537592714001595
27. E. Saez, S. Stancheva, Generalized social marginal welfare weights for optimal tax theory. *National Bureau of Economic Research* **18835** (2013); available at www.nber.org/papers/w18835.
28. I. Kuziemko, M. I. Norton, E. Saez, S. Stancheva, How elastic are preferences for redistribution? Evidence from randomized survey experiments. *National Bureau of Economic Research* **18865** (2013); available at www.nber.org/papers/w18865.
29. P. K. Piff, Wealth and the inflated self: Class, entitlement, and narcissism. *Pers. Soc. Psychol. Bull.* **40**, 34–43 (2014). doi: 10.1177/0146167213501699; pmid: 23963971
30. S. Choi, R. Fisman, D. Gale, S. Kariv, Revealing preferences graphically: An old method gets a new toolkit. *Am. Econ. Rev.* **97**, 153–158 (2007). doi: 10.1257/aer.97.2.153
31. S. G. Bronars, The power of nonparametric tests of preference maximization. *Econometrica* **55**, 693–698 (1987). doi: 10.2307/1913608
32. G. S. Becker, Irrational behavior and economic theory. *J. Polit. Econ.* **70**, 1–13 (1962). doi: 10.1086/258584

ACKNOWLEDGMENTS

We are grateful to J. Andreoni, D. Bernheim, S. DellaVigna, C. Jolls, D. Kahan, B. Polak, M. Rabin, E. Saez and A. Schwartz for helpful discussions and comments. This paper has also benefitted from suggestions by the participants of seminars at several universities and conferences. The experiments reported in this paper were funded by the Columbia University Graduate School of Business, the Center for Equitable Growth at UCB, YLS, and the Oscar M. Reubhausen Fund. S.K. acknowledges the National Science Foundation and the Multidisciplinary University Research Initiative (MURI) for financial support. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any funding agency. Replication data files are available online at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FJYHMRA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/aab0096/suppl/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 and S2
References (*33*–*36*)

16 March 2015; accepted 10 August 2015
10.1126/science.aab0096

RESEARCH ARTICLE

VACCINES

A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen

Antonietta Impagliazzo,[1]*† Fin Milder,[1]‡§ Harmjan Kuipers,[1]‡§ Michelle V. Wagner,[2]‡|| Xueyong Zhu,[3]‡ Ryan M. B. Hoffman,[3]‡ Ruud van Meersbergen,[1]§ Jeroen Huizingh,[1]§ Patrick Wanningen,[1]§ Johan Verspuij,[1]§ Martijn de Man,[1]§ Zhaoqing Ding,[2]|| Adrian Apetri,[1]† Başak Kükrer,[1]† Eveline Sneekes-Vriese,[1] Danuta Tomkiewicz,[1]† Nick S. Laursen,[3]¶ Peter S. Lee,[3] Anna Zakrzewska,[1]§ Liesbeth Dekking,[1]§ Jeroen Tolboom,[1]§ Lisanne Tettero,[1]§ Sander van Meerten,[1]§ Wenli Yu,[3] Wouter Koudstaal,[1]† Jaap Goudsmit,[1]† Andrew B. Ward,[3] Wim Meijberg,[1]§ Ian A. Wilson,[3]* Katarina Radošević[1]#

The identification of human broadly neutralizing antibodies (bnAbs) targeting the hemagglutinin (HA) stem revitalized hopes of developing a universal influenza vaccine. Using a rational design and library approach, we engineered stable HA stem antigens ("mini-HAs") based on an H1 subtype sequence. Our most advanced candidate exhibits structural and bnAb binding properties comparable to those of full-length HA, completely protects mice in lethal heterologous and heterosubtypic challenge models, and reduces fever after sublethal challenge in cynomolgus monkeys. Antibodies elicited by this mini-HA in mice and nonhuman primates bound a wide range of HAs, competed with human bnAbs for HA stem binding, neutralized H5N1 viruses, and mediated antibody-dependent effector activity. These results represent a proof of concept for the design of HA stem mimics that elicit bnAbs against influenza A group 1 viruses.

The ultimate goal of influenza vaccinology is the development of a universal vaccine that protects against a wide range of strains and subtypes, thereby eliminating the need for seasonal reformulation of vaccines and providing an effective defense against viruses with pandemic potential. The relatively recent discovery of broadly neutralizing human monoclonal antibodies (bnAbs) against influenza viruses (*1*–*8*) has raised hopes that a broadly protective vaccine may indeed be feasible (*9*–*13*). Because the majority of these bnAbs are directed toward highly conserved conformational epitopes in the hemagglutinin (HA) stem (*1*, *2*, *4*, *5*, *7*), this region may have the potential to induce broad protective immunity, provided it is well exposed and properly presented.

Various strategies to enhance exposure of the HA stem to the immune system are being explored, including presentation on self-assembling nanoparticles (*14*), chimeric HAs (*15*, *16*), epitope transplantation on a virus-like particle (*17*), and immune refocusing (*18*). Yet another approach involves removal of the HA head while stabilizing the HA stem. A prerequisite for generating a broadly protective soluble HA stem immunogen is that it is stable and structurally resembles the stem of full-length (FL) HA. This simple and elegant concept is complicated by serious structural constraints (*19*). HA is a metastable trimeric surface glycoprotein (*20*, *21*) that undergoes extensive conformational rearrangements at low pH (*22*). Removal of the transmembrane domain and HA head without extensive structural modifications to stabilize the remaining molecule inevitably leads to loss of native conformation of the HA stem (*19*) and concomitant loss of conformational bnAb epitopes. Previously reported soluble HA stem–derived immunogens have exhibited lower affinity with bnAbs relative to FL HA, indicating suboptimal conformation (*23*–*26*).

Design and initial characterization

Starting with the HA sequence of H1N1 A/Brisbane/59/2007, changes were introduced using a combination of rational design and library approaches (*27*). In the first design phase (stages I to III; Fig. 1A), we aimed to generate a soluble and stable HA stem monomer containing bnAb epitopes in the correct conformation, whereas in the final stages (stages IV and V), we aimed to create a mini-HA in its natural trimeric form. Details on the design, sequences, and screening are outlined in (*27*) and figs. S1 and S2. As a readout for a native-like stem conformation of the mini-HA constructs, we used the binding of bnAbs CR9114 and CR6261 (both of which recognize epitopes in the HA stem) (*28*). From each design stage, the best-performing candidates were selected for further optimization. Key mutations in these lead candidates are schematically depicted in fig. S2 and on a model structure of mini-HA in Fig. 1B.

Selected mini-HAs #4157 (stage II), #4454 (stage III), #4650 (stage IV), and #4900 (stage V) were expressed in human embryonic kidney (HEK) 293F cells and purified. On the basis of size, mini-HAs #4157 and #4454 appeared predominantly monomeric in solution; #4650 was larger, suggesting at least partial formation of dimers, whereas #4900 appeared trimeric (fig. S3A). Formation of an intermonomer cysteine bridge in trimeric mini-HA #4900 was confirmed using SDS–polyacrylamide gel electrophoresis under nonreducing and reducing conditions (fig. S3B). All selected mini-HA constructs have four putative N-linked glycosylation sites (Fig. 1B). Under reducing conditions, the mini-HA bands were diffuse and ran at higher molecular weight (35 to 45 kD) than expected (29 kD), indicating that the mini-HAs were glycosylated (fig. S3B). Accordingly, deglycosylation with peptide-N-glycosidase F resulted in sharper bands (~30 kD, as for mini-HA #4900 in fig. S3B). The glycans appear relevant for proper folding, as attempts to efficiently produce mini-HA with retention of bnAb epitopes in bacteria were not successful (*29*).

Size exclusion chromatography with inline multi-angle light scattering (SEC-MALS) analysis revealed a single peak (fig. S4A), with average molecular weights of 40 kD for #4157 and #4454, 60 kD for #4650, and 108 kD for #4900 (Table 1), which also indicated that #4157 and #4454 are monomeric in solution, #4650 forms dimers (at least in part), and #4900 forms trimers. In the presence of CR9114 or CR6261 Fabs, a peak shift to shorter retention time was observed for all mini-HAs, indicating formation of complexes (fig. S4A), which was confirmed by MALS (Table 1), with masses in agreement with binding of one Fab (monomeric mini-HAs #4157 and #4454), one to two Fabs (dimeric mini-HA #4650), or three Fabs (trimeric mini-HA #4900) (*30*).

Binding of CR9114 and CR6261 to all four candidate immunogens was in the nanomolar range, with affinity improving as the design process progressed (Table 1 and fig. S4B). The strongest binding was exhibited by mini-HA #4900, with dissociation constant (K_D) values comparable to those of trimeric FL HA, indicating that this mini-HA recapitulates the bnAb epitopes and thus

[1]Crucell Vaccine Institute, Janssen Center of Excellence for Immunoprophylaxis, Archimedesweg 4-6, 2301 CA Leiden, Netherlands. [2]Crucell Vaccine Institute, Janssen Center of Excellence for Immunoprophylaxis, 3210 Merryfield Row, San Diego, CA 92121, USA. [3]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. ***Corresponding author. E-mail: aimpagli@its.jnj.com (A.I.); wilson@scripps.edu (I.A.W.)** †Present address: Janssen Prevention Center, Janssen Pharmaceutical Companies of Johnson & Johnson, Archimedesweg 4-6, 2301 CA Leiden, Netherlands. ‡These authors contributed equally to this work. §Present address: Infectious Diseases and Vaccines Therapeutic Area, Janssen Research and Development, Janssen Pharmaceutical Companies of Johnson & Johnson, Archimedesweg 4-6, 2301 CA Leiden, Netherlands. ||Present address: Janssen Prevention Center, Janssen Pharmaceutical Companies of Johnson & Johnson, 3210 Merryfield Row, San Diego, CA 92121, USA. ¶Present address: Department of Molecular Biology and Genetics, Aarhus University, Aarhus 8000, Denmark. #Present address: Global Biotherapeutics, Sanofi, 94400 Vitry-sur-Seine, France.

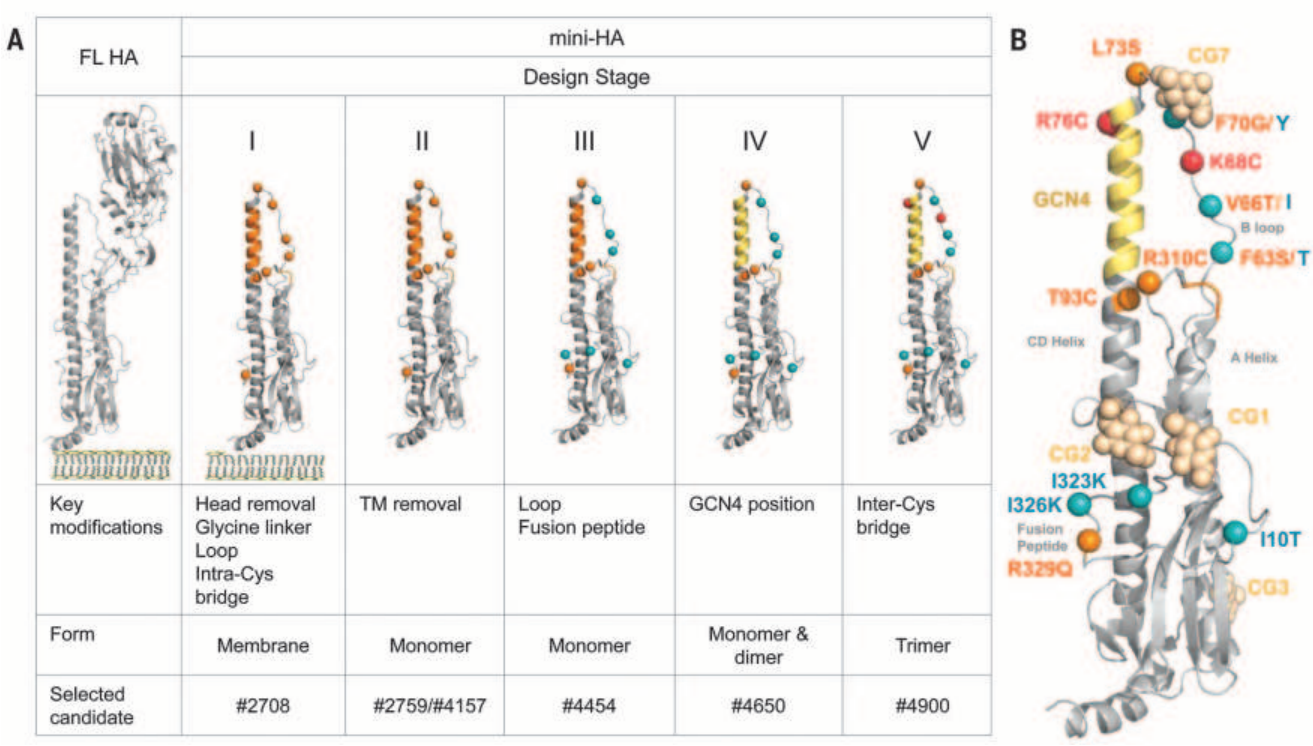


Fig. 1. The mini-HA design strategy and key candidates. (**A**) The mini-HA design strategy is described in five stages, each characterized by specific modifications leading to selection of the best construct for each stage. Key modifications for each stage are schematically depicted and color-coded. Mini-HA #4157 was also generated in an alternative, longer form as mini-HA #2759 with 11 additional C-terminal residues (fig. S1F). (**B**) Ribbon mini-HA model (from PDB ID 1RD8) with color-coded modifications (per stage, as in Fig. 1A) and putative N-glycosylation sites (CG1-3, CG7; taupe).

Table 1. Binding of bnAbs to selected mini-HA candidates. bnAbs CR9114 and CR6261 form high-affinity complexes with mini-HAs. K_D values were determined using biolayer interferometry and steady-state analysis. MALS analysis and size determination were used to confirm Fab binding and complex formation. The numbers are means and SDs of at least two independently produced protein batches. The values in parentheses represent the expected molecular weights for monomeric (#4157, #4454), dimeric (#4650), and trimeric (#4900) mini-HA–Fab complexes, using the determined molecular weight of the mini-HA and 48 kD for the Fabs.

Construct	K_D (nM)		Molecular weight (kD)			
	CR9114	CR6261	Mini-HA	Mini-HA with Fab CR9114	Mini-HA with Fab CR6261	Mini-HA with Fab CR8020
Mini-HA						
#4157	28.3 ± 7.4	490 ± 192	40.3 ± 1.2	78.3 ± 0.6 (88)	63.7 ± 3.2 (88)	42.3 ± 2.1
#4454	9.5 ± 0.7	105 ± 36	40.0 ± 1.4	81.5 ± 2.1 (88)	72.5 ± 2.1 (88)	40.5 ± 2.1
#4650	2.7 ± 0.6	4.5 ± 1.3	59.5 ± 2.1	129.5 ± 2.1 (156)	109.5 ± 0.7 (156)	62.0 ± 1.4
#4900	0.50 ± 0.04	0.6 ± 0.1	107.5 ± 0.7	241.5 ± 0.7 (252)	214.5 ± 2.1 (252)	110.5 ± 0.7
FL HA						
Monomer	6.7 ± 4.0	11.7 ± 4.2				
Trimer	0.8 ± 0.3	0.6 ± 0.2				

accurately mimics the native conformation of the HA stem.

Structural characterization

Modifications introduced throughout the design gradually increased the conformational stability as determined by hydrogen-deuterium exchange mass spectrometry (HDX-MS) (*31*, *32*). In particular, modifications introduced in stage III (from #4157 to #4454) and stage V (from #4650 to #4900) stabilized the overall structure and the shorter A helix (Fig. 2). Furthermore, we observed only a low level of deuterium exchange on the long CD helix of mini-HA #4900, likely due to the shielding effect caused by trimerization. Most important, the HDX pattern of mini-HA #4900 is very similar to that of trimeric FL HA, further demonstrating that the final design mimics the structure and dynamics of native HA. In line with these findings, the overall stability of the mini-HAs also evolved, as is apparent from measurements of the thermal stability (fig. S5).

Complexes of mini-HAs with CR9114 Fab were first studied by negative-stain electron microscopy (EM) (Fig. 3A). The EM class averages illustrate the progression from monomeric (#4454) to a more heterogeneous mixture of monomers and dimers (#4650) and finally to a trimeric form (#4900). To provide further structural insights, we determined the crystal structures of #4454-CR9114 and #4900-CR9114 Fab complexes to 4.3 Å and 3.6 Å resolution, respectively (Fig. 3, B and C, and table S1) (*27*). In the #4454-CR9114 complex (Fig. 3B), CR9114 binds monomeric mini-HA and recognizes essentially the same epitope as in FL HA through contacts only with its heavy chain (table S2) (*11*). The solvent-accessible surfaces of the CR9114 epitope residues in #4900 are similar to those in A/California/04/2009 FL H1 HA (table S3). The #4900 trimer differs only slightly from the stem of trimeric A/California/04/2009 FL HA (and perhaps other HAs), as it adopts a more open, splayed conformation at the trimer base with the HA2 H helix, shifted by about 14.5 Å (Fig. 3, C and D), perhaps as a consequence of insertion of the GCN4 trimerization motif at the top of the long helix in #4900 (fig. S6). The primary amino acid interactions are maintained between CR9114 and #4454 as compared to those between CR9114 and FL H5 HA (*3*), and the CR9114 complex is trimeric

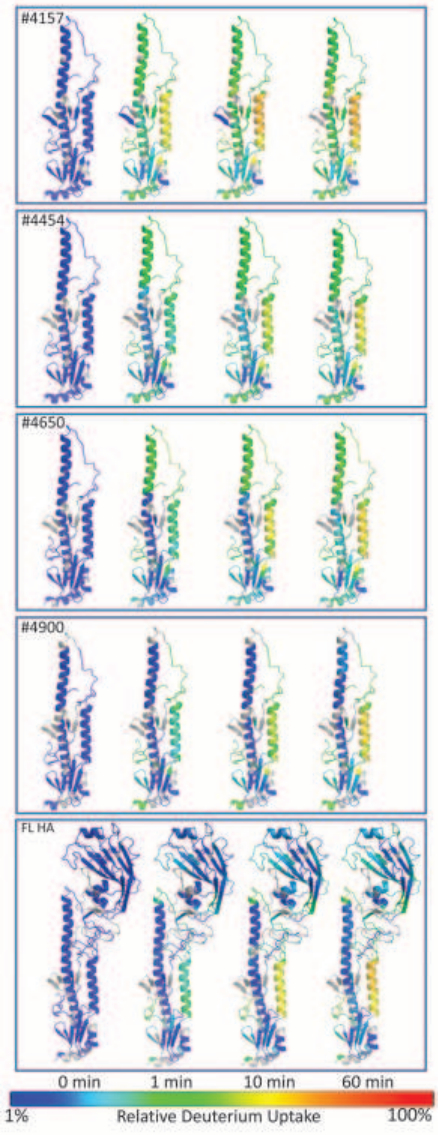


Fig. 2. Dynamic structural properties of mini-HAs and trimeric FL HA analyzed by hydrogen-deuterium exchange mass spectrometry. Relative deuterium uptake over time (0 to 60 min), illustrating the local dynamics of different structural units, is depicted on model mini-HA and FL HA structures using a color code from blue (1% deuterium uptake; highly protected) to red (100% deuterium uptake; highly exposed or flexible).

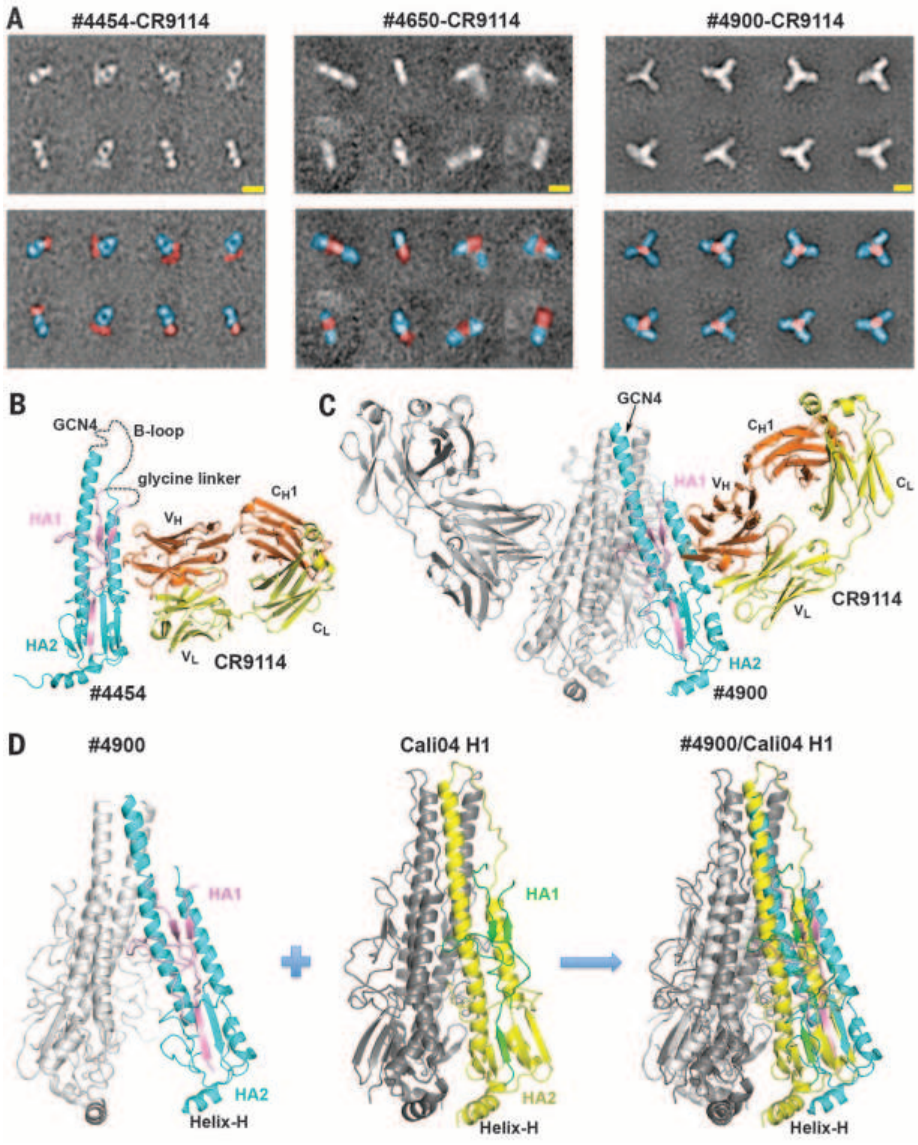


Fig. 3. EM and crystal structures of mini-HAs in complex with CR9114 Fab. (**A**) Representative reference-free class averages of mini-HAs #4454, #4650, and #4900 in complex with CR9114 Fab using negative-stain EM. The colored panels indicate how densities are attributed to either mini-HA (red) or Fab (blue). The yellow scale bar shown is 36.5 Å for the #4454 and #4650 complexes, and 73.0 Å for the #4900 complex. (**B**) Crystal structure of monomeric mini-HA #4454 (pink, HA1; cyan, HA2) in complex with CR9114 Fab (yellow, light chain; orange, heavy chain). The GCN4 motif, the B-loop, and the glycine linker are shown in dashed lines because they are disordered in the structure. (**C**) Crystal structure of trimeric #4900 in complex with CR9114 Fab. One monomer of #4900-CR9114 Fab trimer is shown and color-coded as in (B); the other two monomers of #4900-CR9114 Fab trimer are shown in gray for clarity. The GCN4 motif at the top of the long helix in #4900 is ordered and can be modeled. (**D**) Superimposition of mini-HA #4900 with a "mini-HA" model computationally extracted from the Cali04 HA structure [HA from A/California/04/2009 (H1N1), PDB ID 4M4Y] (one monomer with HA1 in green and HA2 in yellow, and the other two monomers in dark gray). See also tables S2 and S3.

(Fig. 3C). The #4900 protomer is superimposable with an HA protomer from the H1 A/California/04/2009 trimer (PDB ID 4M4Y) with a C_α RMSD of 1.2 Å, and only 1.0 Å for the CR9114 epitope residues. As a consequence of the removal of the HA1 head in #4900 and a more open trimer base (table S2), some residues in the HA2 A helix, B loop, C and D helix, and E strand become solvent-accessible. These results convincingly demonstrate that the iterative design process led to antigens with an increasingly improved HA stem conformation and ultimately yielded HA-like stem antigens with a preserved CR9114 epitope.

In vivo effects

When tested in mice, all mini-HAs were highly immunogenic, eliciting high titers of antibodies binding to the FL HA of the A/Brisbane/59/2007 (H1N1) strain used as the basis for mini-HA generation (fig. S7A), as well as to FL HAs from a number of other group 1 (H1, H5, H9) and group 2 (H3, H7) influenza A strains (fig. S7B). These results demonstrate that antibodies induced with mini-HAs efficiently recognize epitopes that are present in native sequences of FL HA and are

conserved among different group 1 and even some group 2 influenza A strains (*33*).

Next, we evaluated the ability of mini-HA immunogens to provide protection against stringent lethal challenges and characterized the elicited immune response (Fig. 4 and fig. S7). Protection against influenza challenge with H1N1 A/PR/8/34 (heterologous to the parental strain used for design of the mini-HAs) was assessed after one, two, or three immunizations with mini-HAs #2759 (monomer, long variant of #4157; fig. S2), #4650 (mix of monomers and dimers), and #4900 (trimer). Mice immunized three times with monomeric mini-HA #2759 were only partially protected against mortality and showed severe weight loss and clinical symptoms, whereas #4650 and #4900 provided complete protection, with #4900-immunized mice showing neither weight loss nor clinical symptoms (Fig. 4A and fig. S7C). Only mini-HA #4900 showed the same full protective ability when administered twice. Even after one immunization, #4900 exhibited 90% protection against mortality, although animals exhibited some (recoverable) weight loss and clinical symptoms, whereas #4650 and #2759 were not able to protect animals. Antibodies elicited with mini-HA #4900 after two and three immunizations competed with bnAb CR9114 for binding to the FL HA stem (Fig. 4A), which shows that this mini-HA could readily elicit antibodies specific for bnAb stem epitopes.

Mice immunized three times with an expanded set of mini-HAs were challenged with heterosubtypic H5N1 A/Hong Kong/156/97 virus to evaluate the breadth of protection. Mini-HAs #2759 and #4157 (monomers) failed to elicit protection, #4454 (improved library monomer) and #4650 exhibited partial protection, whereas #4900 exhibited complete protection with neither weight loss nor clinical symptoms (Fig. 4B and fig. S7D). To further investigate antibody-mediated effector mechanisms, we tested prechallenge sera in a neutralization assay using pseudoparticles derived from H5N1 A/Vietnam/1194/04 (*34*). Sera from mice immunized with mini-HA #4900, and to some extent #4650, showed detectable neutralization, demonstrating the ability of these mini-HAs to elicit bnAbs (Fig. 4C). Besides direct virus neutralization, Fc-mediated effector mechanisms, such as antibody-dependent cellular cytotoxicity (ADCC), contribute substantially to protection against influenza, with stem-directed bnAbs being particularly effective in these mechanisms (*35*). We therefore used a mouse-adapted ADCC surrogate assay to test prechallenge sera (*36*–*38*). The evolving mini-HAs demonstrated an increasing ability to elicit antibodies with potential to induce ADCC (Fig. 4D), with ADCC potency of serum antibodies corresponding to survival proportions observed in the H5N1 challenge model. In accordance with the binding data (fig. S7, A and B), mini-HA #4900 elicited ADCC responses to a wide range of group 1 influenza A HAs (fig. S7E). To assess whether serum antibodies are indeed responsible for the observed in vivo protection, we passively transferred serum from mice immunized with #4650 and #4900 to naïve mice and subsequently challenged the recipient mice with A/HK/156/97 (H5N1) virus (Fig. 4E and fig. S7F). In contrast to mice that received mock serum, mice that received serum from mini-HA–immunized mice were significantly protected, proving that antibodies play a key role in the protection elicited by mini-HAs.

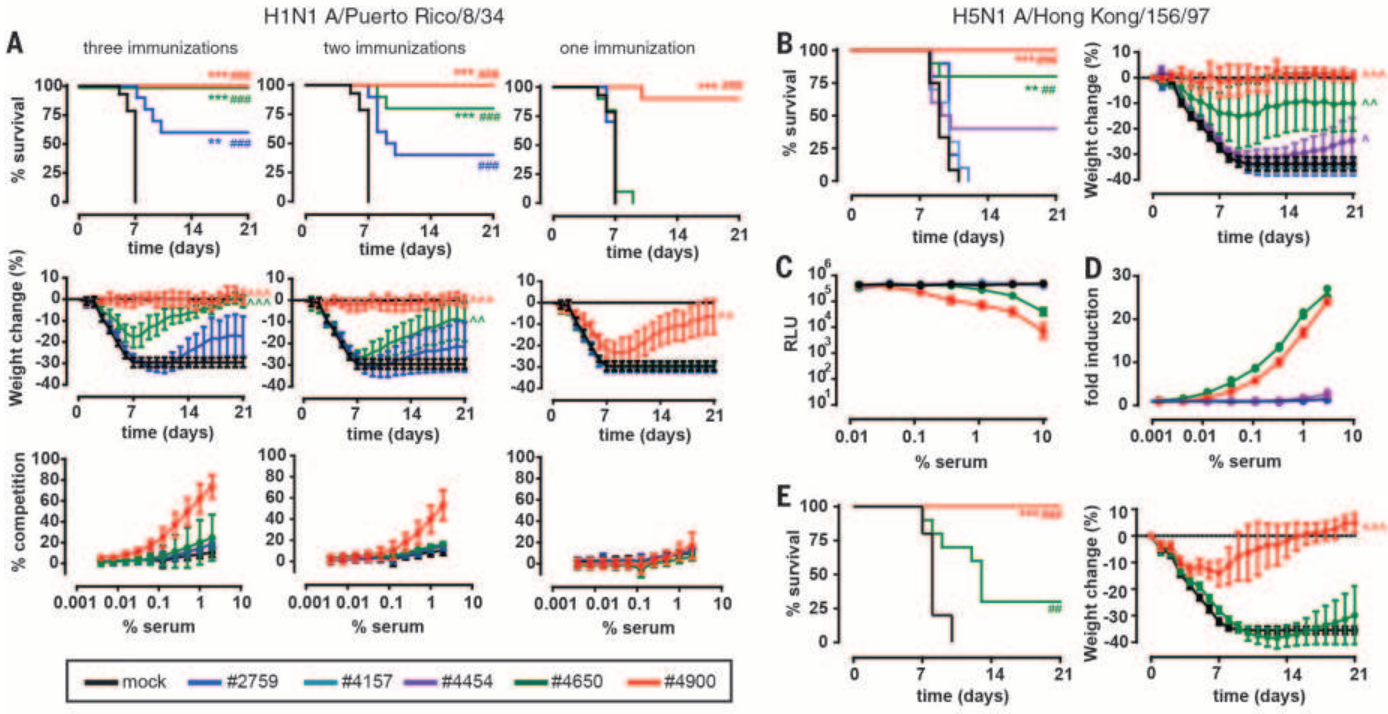


Fig. 4. Immunogenicity and protective efficacy of mini-HA candidates in mice. (**A**) Kaplan-Meier survival curves (top panels) and mean body weight change (middle panels) after one, two, or three immunizations with mini-HA ($n = 10$) or vehicle (mock; $n = 14$), followed 4 weeks later by challenge with 25 times the lethal dose for 50% of mice (LD_{50}) H1N1 A/PR/8/34. Serially diluted prechallenge sera were individually tested for competition with bnAb CR9114 (bottom panels). Error bars denote SD of the mean for CR9114 competition graphs. (**B**) Kaplan-Meier survival curves and mean body weight change of animals ($n = 10$ for mini-HA groups, $n = 12$ for mock group) immunized three times and challenged with 12.5 LD_{50} H5N1 A/Hong Kong/156/97. (**C**) Neutralization of H5 HA A/Vietnam/1194/04 pseudovirus, with prechallenge serum pooled per immunization group. Data shown are geometric means ± SD of three technical replicates per measurement. Results are representative of three independent experiments. (**D**) ADCC surrogate assay with target cells expressing H5N1 A/Hong Kong/156/97 HA and prechallenge serum pooled per immunization group. Individual data points and the line representing the geometric mean of two technical replicates per measurement are shown, representative of two independent experiments. (**E**) Kaplan-Meier survival curves and mean body weight change for animals receiving serum from animals immunized three times with mini-HA or mock ($n = 10$ per group). Four weeks after the last immunization, serum was pooled and transferred to naïve recipient mice on three consecutive days before challenge, followed by challenge with 12.5 LD_{50} H5N1 A/Hong Kong/156/97. Symbols indicate significantly improved survival proportion (Fisher's exact test; *), survival time (log-rank test; #) or body weight change (analysis of variance; ^) relative to mock-immunized animals. Three symbols (i.e., ***, ###, ^^^) indicate $P < 0.001$, two symbols $P < 0.01$, and one symbol $P < 0.05$.

These combined results demonstrate that the progressive structural refinement and stabilization of mini-HAs directly translates to increasing ability to elicit stem-targeting, neutralizing, and ADCC-mediating antibodies that can protect mice against heterologous and heterosubtypic group 1 influenza strains, with trimeric #4900 exhibiting the most favorable features. In a recent study (*26*), a bacterially produced trimeric HA stem construct also seemed more effective than monomeric versions in eliciting protection in an animal influenza challenge model.

We next tested the immunogenicity and protective efficacy of a three-dose regimen of mini-HA #4900 in six cynomolgus monkeys (*Macaca fascicularis*). As controls, groups of six monkeys received three injections of phosphate-buffered saline (PBS) or two human doses of a trivalent seasonal influenza vaccine (the standard of care for naïve children). As previously observed in mice, #4900 elicited high titers of antibodies in monkeys that were able to bind to a wide range of group 1 FL HAs (Fig. 5A and fig. S8A); compete with stem-binding bnAb CR9114 for binding to homologous, heterologous, and heterosubtypic HAs (Fig. 5B); and elicit ADCC responses to these divergent HAs (Fig. 5C). Furthermore, #4900 elicited H5N1 neutralizing antibodies that were detectable not only in a pseudoparticle-based neutralization assay (fig. S8B), as for the mouse sera (Fig. 4C), but also in a standard microneutralization assay (Fig. 5D). Relative to mock-immunized monkeys, animals vaccinated with mini-HA #4900 had significantly reduced fever after challenge with A/Mexico/InDRE4487/2009 (H1N1) virus (Fig. 5E and fig. S8E) (*39*), although no significant effect on the tracheal viral load was detected (fig. S8C). These results are comparable to those from monkeys vaccinated with the trivalent seasonal influenza vaccine (Fig. 5E and fig. S8C). However, the immune response elicited by mini-HA #4900 differed from the response elicited with the trivalent vaccine. As expected, vaccination with the seasonal vaccine, but not with "headless" #4900, elicited hemagglutination-inhibiting (HI) antibodies against A/California/07/09 (H1N1) virus, from which the H1 component of the trivalent vaccine is derived and which is closely related (two amino acid differences) to the A/Mexico/InDRE4487/2009 (H1N1) virus used to challenge the monkeys (fig. S8D). Furthermore, because the A/Texas/50/2012 HA antigen was present in the trivalent seasonal vaccine, the vaccine elicited higher antibody titers against this H3 strain than did mini-HA #4900 (fig. S8A). However, unlike the mini-HA–immunized monkeys, no antibodies able to compete with CR9114, induce ADCC, or neutralize H5N1 virus were detectable in the sera of monkeys immunized with the seasonal vaccine (Fig. 5, B to E, and fig. S8B); these results show that the mini-HA elicited a qualitatively different and much broader immune response than the seasonal vaccine.

In these in vivo studies, we used a single antigen dose and a potent adjuvant. Although this adjuvant has not yet been registered for human use, it has been tested in a number of clinical

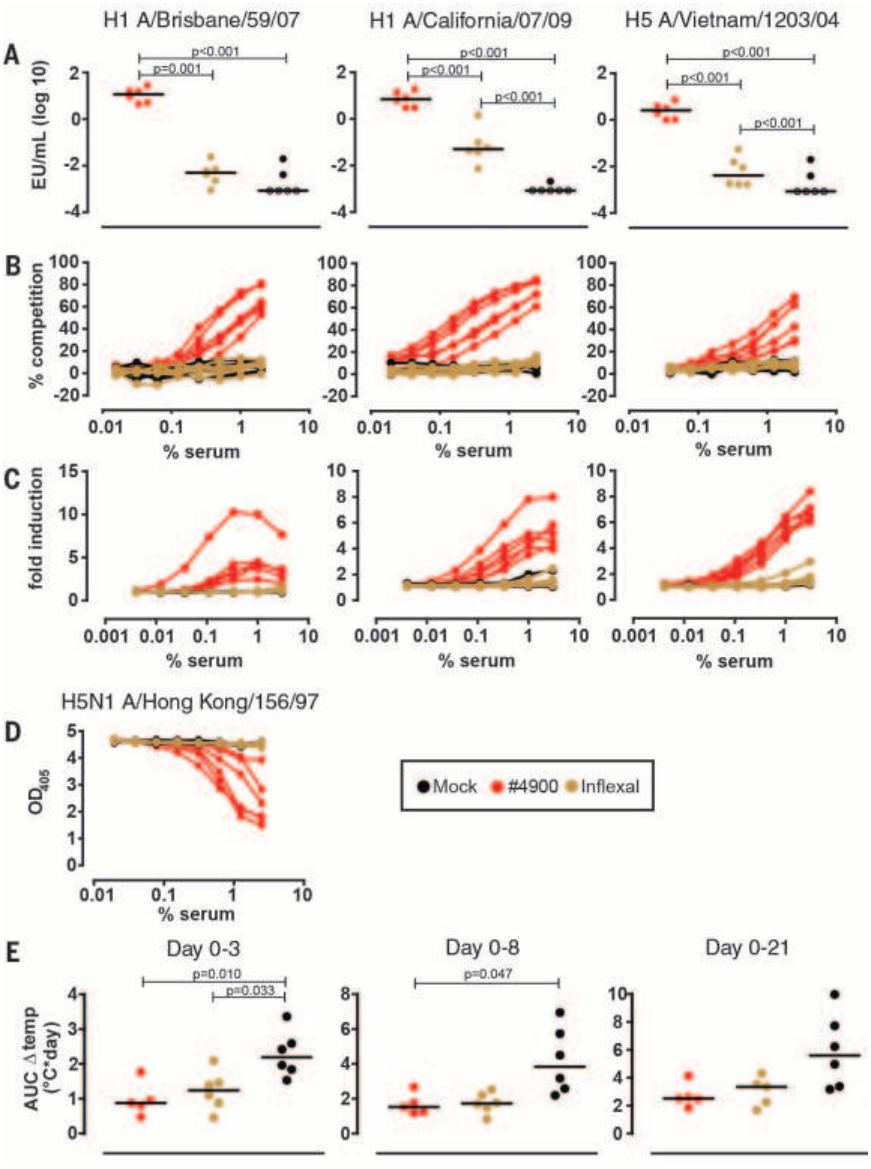


Fig. 5. Immunogenicity and protective efficacy of mini-HA #4900 in cynomolgus monkeys. Animals (n = 6 per group) were immunized with mini-HA #4900, Inflexal V [a trivalent virosomal seasonal influenza subunit vaccine of the 2013 season (H1N1 A/California/07/2009, H3N2 A/Texas/50/2012, and B/Massachusetts/2/2012; 15 μg of HA per strain per vaccine dose], or vehicle (mock) followed by challenge with 4×10^6 median tissue culture infectious dose of H1N1 A/Mexico/InDRE4487/09. (**A**) Prechallenge serum antibody titers to HA from H1N1 A/Brisbane/59/07 (left), H1N1 A/California/07/09 (center), and H5N1 A/Vietnam/1203/04 (right). Open symbols indicate samples that were below detection limits using a serum start dilution of 1/50. Statistical analysis between treatments was performed using pairwise t test with Tukey-Kramer adjustment for multiple comparisons. Bars represent group medians. (**B**) Serially diluted prechallenge sera from individual animals tested for competition with bnAb CR9114, using FL HA from H1N1 A/Brisbane/59/07 (left), H1N1 A/California/07/09 (center), and H5N1 A/Vietnam/1203/04 (right) as antigen. Data shown are the mean of two technical replicates per measurement. (**C**) ADCC surrogate assay with target cells expressing HA from H1N1 A/Brisbane/59/07 (left), H1N1 A/California/07/09 (center), and H5N1 A/Vietnam/1203/04 (right) and prechallenge serum from individual animals. Data shown are the geometric mean of two technical replicates per measurement. (**D**) Neutralization of H5N1 A/Hong Kong/156/97 reassortant virus using serially diluted prechallenge serum from individual animals. Data shown are the mean of two technical replicates per measurement. (**E**) Cumulative net body temperature increase after challenge per animal. The area under the curve of the net temperature increase was calculated over intervals of day 0 to 3 (left), day 0 to 8 (center), and day 0 to 21 (right). Statistical analysis between treatments was performed using pairwise t test with Tukey-Kramer adjustment for multiple comparisons. Bars denote median. One animal of the #4900 group was excluded because of data logger failure, and one animal of the Inflexal group died at the end of day 8 and was excluded from the day 0 to 21 interval calculation.

trials with favorable safety and immunogenicity outcomes (*40*). Future studies will need to address the minimal doses of mini-HA and adjuvant for a protective response. Because preexisting immunity may have a profound effect on the breadth of the response and no animal model recapitulates the complex preexisting immunity against influenza found in humans, such studies should be performed in humans (*41*).

Concluding remarks

We have described the design and characterization of a series of soluble HA immunogens solely composed of the HA stem. Although all selected mini-HAs elicited comparable levels of antibodies to FL HA, the breadth and protective ability of the elicited antibodies progressively increased with the structural evolution of mini-HA configuration. The final candidate—stabilized trimeric mini-HA #4900—demonstrated its unique ability to elicit broad and protective immune response in mice and nonhuman primates. It has been reported (*42*, *43*) that stabilization of respiratory syncytial virus F antigen improves immune response and protection. Our results demonstrate that the same principle holds for influenza HA and provide further direction for the design of an epitope-based, universal influenza vaccine.

REFERENCES AND NOTES

1. D. Corti *et al.*, *Science* **333**, 850–856 (2011).
2. M. Throsby *et al.*, *PLOS ONE* **3**, e3942 (2008).
3. C. Dreyfus *et al.*, *Science* **337**, 1343–1348 (2012).
4. D. C. Ekiert *et al.*, *Science* **324**, 246–251 (2009).
5. D. C. Ekiert *et al.*, *Science* **333**, 843–850 (2011).
6. R. H. Friesen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 445–450 (2014).
7. J. Sui *et al.*, *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
8. A. Kashyap *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5986–5991 (2008).
9. D. R. Burton, P. Poignard, R. L. Stanfield, I. A. Wilson, *Science* **337**, 183–186 (2012).
10. G. L. Chen, K. Subbarao, *Nat. Med.* **15**, 1251–1252 (2009).
11. D. C. Ekiert, I. A. Wilson, *Curr. Opin. Virol.* **2**, 134–141 (2012).
12. T. T. Wang, P. Palese, *Nat. Struct. Mol. Biol.* **16**, 233–234 (2009).
13. T. T. Wang, P. Palese, *Science* **333**, 834–835 (2011).
14. M. Kanekiyo *et al.*, *Nature* **499**, 102–106 (2013).
15. R. Hai *et al.*, *J. Virol.* **86**, 5774–5781 (2012).
16. N. Pica *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2573–2578 (2012).
17. A. Schneemann *et al.*, *J. Virol.* **86**, 11686–11697 (2012).
18. D. Eggink, P. H. Goff, P. Palese, *J. Virol.* **88**, 699–704 (2014).
19. J. Cohen, *Science* **341**, 1171 (2013).
20. J. J. Skehel, M. D. Waterfield, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 93–97 (1975).
21. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366–373 (1981).
22. J. J. Skehel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 968–972 (1982).
23. G. Bommakanti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13701–13706 (2010).
24. G. Bommakanti *et al.*, *J. Virol.* **86**, 13434–13444 (2012).
25. Y. Lu, J. P. Welsh, J. R. Swartz, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 125–130 (2014).
26. V. V. Mallajosyula *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2514–E2523 (2014).
27. See supplementary materials on *Science* Online.
28. CR9114 and CR6261 share a nearly identical epitope formed by helix A of HA2 and a segment of HA1, and both neutralize virtually all group 1 influenza A strains. CR9114 also neutralizes influenza A group 2 viruses and binds influenza B strains, demonstrating greater breadth relative to CR6261. Because CR6261 appears to have more stringent epitope requirements than CR9114, we used CR6261 binding as the key criterion for the selection of final candidates.
29. The apparent importance of the glycans may partially explain why previously reported HA stem–derived proteins that were expressed in *E. coli*, and thus not glycosylated, failed to mimic the native HA stem more closely.
30. SEC-MALS analysis of mini-HAs in combination with Fab of bnAb CR8020, specific for influenza group 2 viruses, showed neither peak shift nor change in retention time for any of the proteins.
31. L. Konermann, J. Pan, Y. H. Liu, *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
32. H. Wei *et al.*, *Drug Discov. Today* **19**, 95–102 (2014).
33. Structural differences between the stems of group 1 and group 2 HAs restrict the activity of most bnAbs to one group. Although some antibodies, like CR9114, can bind both groups, such antibodies appear to be extremely rare. The observed lower reactivity to group 2 HAs of immune serum elicited against group 1 HA stem mimics was therefore to be expected.
34. I. Alberini *et al.*, *Vaccine* **27**, 5998–6003 (2009).
35. D. J. DiLillo, G. S. Tan, P. Palese, J. V. Ravetch, *Nat. Med.* **20**, 143–151 (2014).
36. Z. J. Cheng *et al.*, *J. Immunol. Methods* **414**, 69–81 (2014).
37. B. S. Parekh *et al.*, *MAbs* **4**, 310–318 (2012).
38. A. Schnueriger *et al.*, *Mol. Immunol.* **48**, 1512–1517 (2011).
39. The reduction in fever in #4900-vaccinated animals relative to PBS-immunized animals is significant when periods shortly after challenge (e.g., 0 to 3 days or 0 to 8 days) are considered, because the fever is most pronounced shortly after challenge. Although still visible when the entire 21-day follow-up period is considered, the reduction is no longer statistically significant because of the large variation in measured body temperatures.
40. R. J. Cox *et al.*, *Vaccine* **29**, 8049–8059 (2011).
41. R. Roozendaal *et al.*, *PLOS ONE* **9**, e103550 (2014).
42. B. E. Correia *et al.*, *Nature* **507**, 201–206 (2014).
43. J. S. McLellan *et al.*, *Science* **342**, 592–598 (2013).

ACKNOWLEDGMENTS

We thank R. Vogels, R. van der Vlugt, D. Zuijdgeest, N. Kroos, V. Klaren, S. Schmit-Tillemans, L. Kil, S. Barens, and O. Diefenbach for scientific input and technical support; J. Klap, M. Koldijk, and G. J. Weverling for statistical support; Algonomics for input in early designs; Novavax for supplying Matrix-M; Promega for early access to mouse surrogate ADCC assay; E. Montomoli, D. Perini, and S. Piccirella from VisMederi for pseudoparticle assay; Janssen Diagnostics for quantitative polymerase chain reaction of tracheal swabs; P. Mooij and G. Koopman of the Biomedical Primate Research Center (BPRC), TNO Triskelion, Central Veterinary Institute, and PreClinBiosystems AG for performing the animal studies; and L. Dolfing and A. Dingemans for critical reading of the manuscript. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. X-ray diffraction data sets were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2. Use of the SSRL, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by National Institute of General Medical Sciences (NIGMS) grants including P41GM103393. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH. Coordinates and structure factors of the crystal structures are deposited in the Protein Data Bank as entries 5CJQ and 5CJS. Crucell Holland B.V., a Janssen company, has the following pending patent applications in this field: WO 2013/079473, WO 2014/191435, WO2008/028946, WO2013/007770, U.S. 62/062,746, and U.S. 62/062,754. Sharing of materials will be subject to standard material transfer agreements.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1301/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References (*44*–*65*)

26 November 2014; accepted 29 July 2015
Published online 24 August 2015
10.1126/science.aac7263

REPORTS

METALLIC GLASSES

Fractal atomic-level percolation in metallic glasses

David Z. Chen,[1]* Crystal Y. Shi,[2]† Qi An,[3]† Qiaoshi Zeng,[4,5] Wendy L. Mao,[2,6] William A. Goddard III,[3] Julia R. Greer[1,7]

Metallic glasses are metallic alloys that exhibit exotic material properties. They may have fractal structures at the atomic level, but a physical mechanism for their organization without ordering has not been identified. We demonstrated a crossover between fractal short-range (<2 atomic diameters) and homogeneous long-range structures using in situ x-ray diffraction, tomography, and molecular dynamics simulations. A specific class of fractal, the percolation cluster, explains the structural details for several metallic-glass compositions. We postulate that atoms percolate in the liquid phase and that the percolating cluster becomes rigid at the glass transition temperature.

Freeze a liquid fast enough, and it becomes a glass, a material that is structurally similar to the liquid but incapable of flow. This concept, albeit not well understood (*1*, *2*), is so ubiquitous that it holds even for metals (*3*). Vitrified metals, or metallic glasses, are a class of disordered materials with nondirectional bonding and possess a suite of lucrative mechanical properties, such as high elastic limit and strength (*4*). Unlike most crystalline metals and alloys, metallic glasses earn their name from a lack of long-range atomic order and the absence of typical defects, such as dislocations, rendering their microstructure challenging to conceptualize

and model. Some studies suggest the existence of short-range order, for which solute-centered clusters serve as the building blocks, and medium-range order characterized by cluster packing (*5*–*7*). These short- and medium-range packing schemes inevitably break down over longer coordinates as a result of spatial incompatibility, and they do not fully describe the atomic organization within these complex glasses. The incomplete understanding of atomic-level structure in glassy materials has made it challenging to capture the physics of their response to mechanical deformation. We propose a model that describes a short-range order and encompasses the long-range structural details of metallic glasses. The model has considerable implications for understanding glass properties and the origin of the glass transition.

Diffraction experiments characterize the structure of amorphous materials by mapping the atomic neighbor–separation distances and statistical density distributions. Dissimilar glasses and liquids commonly possess distinct short- and medium-range orders due to variations in chemical bonding, but the atomic structure becomes fluid like and nearly indistinguishable among different glasses beyond the first few nearest neighbors (*8*). The similarity of atomic-level environments in liquids and glasses makes it difficult to understand how glasses get their rigidity. Glass rigidity may be related to the jamming of atoms as density increases (*9*). The marked difference between the short- and long-range configurations in glassy systems sets glasses apart from crystals. In contrast to crystals, simplifying the underlying structure in a glass is problematic, because the short- and medium-range

[1]Division of Engineering and Applied Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [2]Department of Geological Sciences, Stanford University, Stanford, CA 94305, USA. [3]Materials and Process Simulation Center, California Institute of Technology, Pasadena, CA 91125, USA. [4]Center for High Pressure Science and Technology Advanced Research, Shanghai 201203, China. [5]High Pressure Synergetic Consortium, Geophysical Laboratory, Carnegie Institute of Washington, Argonne, IL 60439, USA. [6]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]Kavli Nanoscience Institute, California Institute of Technology, Pasadena, CA 91125, USA.
***Corresponding author. E-mail: dzchen@caltech.edu**
†These authors contributed equally to this work.

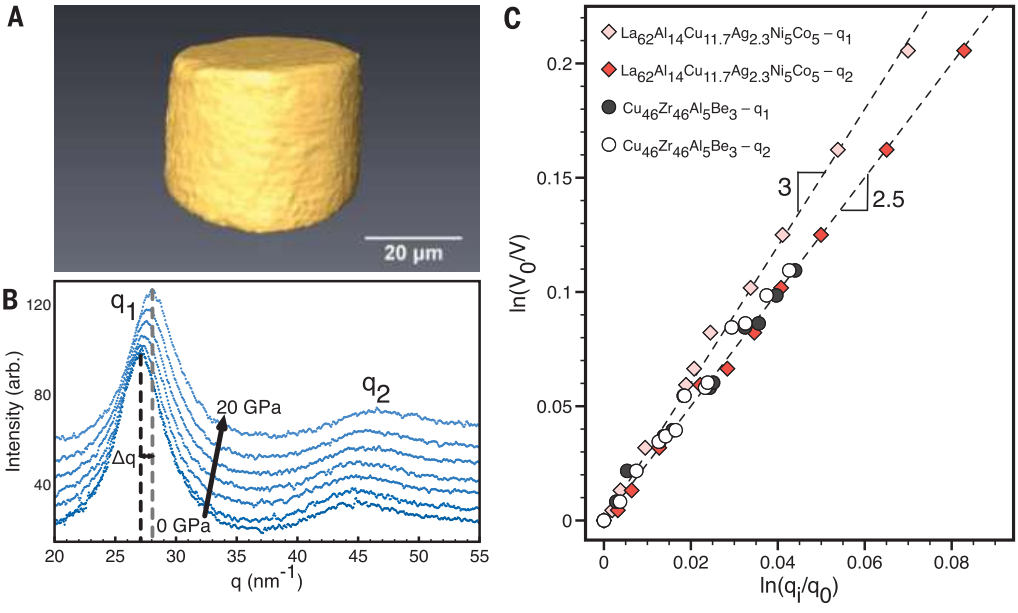


Fig. 1. In situ diffraction and volume results. (**A**) Three-dimensional reconstructed sample volumes from in situ transmission x-ray microscopy data at ~0 GPa. (**B**) In situ x-ray diffraction data with increasing pressure (arb., arbitrary units). (**C**) Volume scaling with scattering vectors $\mathbf{q_1}$ and $\mathbf{q_2}$ for $Cu_{46}Zr_{46}Al_5Be_3$ and $La_{62}Al_{14}Cu_{11.7}Ag_{2.3}Ni_5Co_5$ metallic glasses.

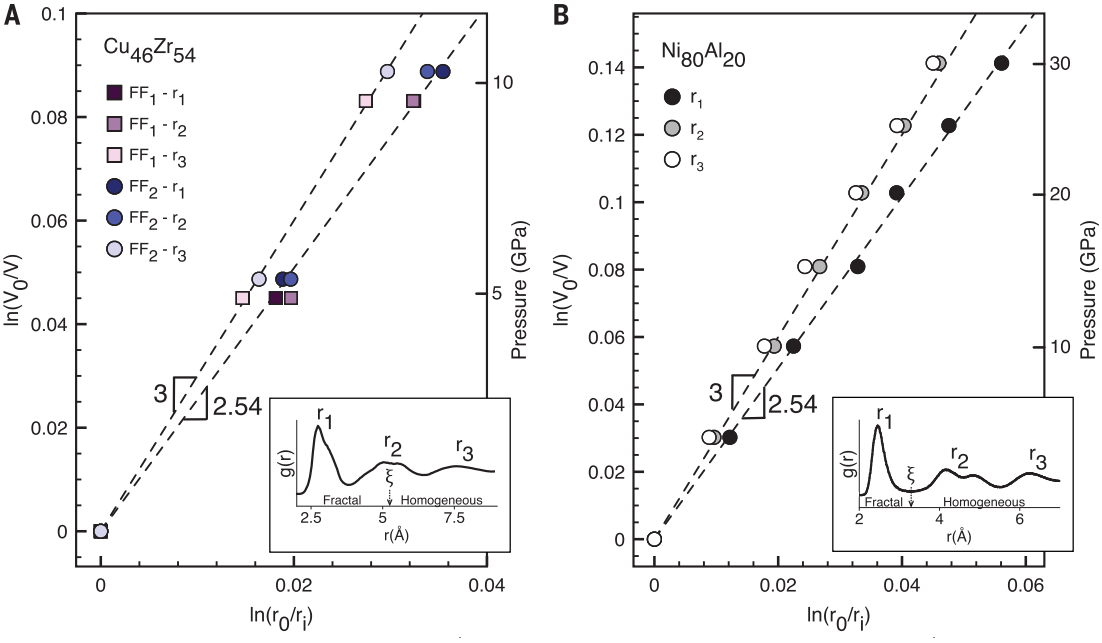


Fig. 2. Dimensionality crossover in simulations. (**A**) The $Cu_{46}Zr_{54}$ from FF_1 and FF_2 both exhibit a transition in dimensionality from ~2.5 to 3 between r_2 and r_3. (**B**) $Ni_{80}Al_{20}$ exhibits a transition in dimensionality from ~2.5 to 3 between r_1 and r_2. The insets show corresponding RDF curves with the correlation lengths ξ indicated.

orders do not repeat in a recognizable pattern. For this reason, no two glasses, produced under the same conditions and with similar diffraction patterns, are identical at the atomic level. The question of how repeatable long-range structures in glasses can emerge from nonrepeating atomic clusters remains unanswered.

Studies suggest fractal properties in metallic glasses (*10*, *11*). Fractal behavior manifests in the relationship between mass (M) and volume. For crystals, this relationship, $M(r) \sim r^3$ (where r is the radius of a region within the material), has a dimensionality (D) of 3. The dimensionality of metallic glasses is closer to ~2.5 (*11*), and any non-integer D corresponds to a fractal (*12*). Many naturally occurring random fractals have $D \sim 2.5$, including crumpled balls of paper and thin sheets (*13*), which are fractals down to the size of nanoballs of graphene oxide (*14*). Fractal concepts may be useful in developing an atomic-level understanding of amorphous materials, because they imply underlying order in inherently chaotic and random arrangements. The specific nature of fractals in metallic glasses is not obvious, because most mass fractals have macroscopic pores at large r (e.g., crumpled paper), and metallic glasses are monolithic materials. Metallic glasses have packing fractions close to or exceeding those of close-packed crystalline metals (*15*). The puzzle of how metallic glasses can simultaneously possess fractal properties and remain fully dense is unresolved (*16*). One possible explanation is that the diffraction experiments only probe the short-range dimensionality. In this work, we observed a fractal short-range $D < 3$ and a homogeneous long-range $D = 3$ for several metallic glasses, indicating the presence of a dimensionality crossover at an intermediate length scale.

Previous studies have focused on the principal (first) diffraction peak only (*10*, *11*). We extended the analysis beyond the first peak, because the information contained in diffraction experiments is spread out in momentum space, and each peak contains information that represents a part of the total structure. We conducted in situ high-pressure x-ray diffraction and full-field nanoscale transmission x-ray microscopy experiments on ~40-μm-diameter cylindrical samples of $Cu_{46}Zr_{46}Al_5Be_3$ metallic glass (Fig. 1A). We made diffraction and sample volume measurements in situ as a function of hydrostatic pressure in a diamond anvil cell. We related the scattering vector (**q**) from diffraction peak positions to volume by increasing the hydrostatic pressure from ~0 to 20 GPa (Fig. 1B). Compared with

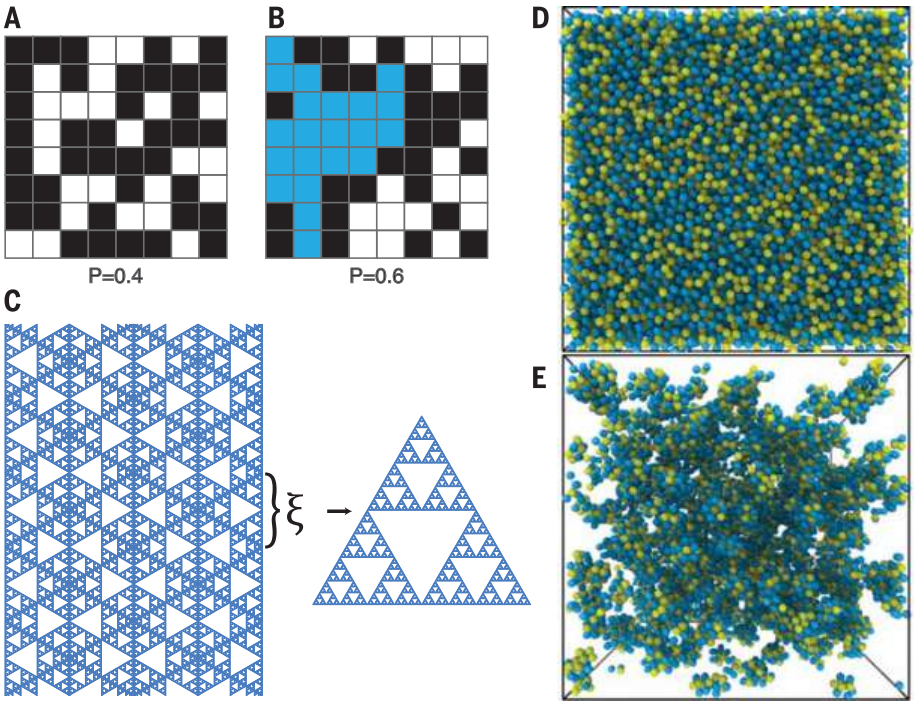


Fig. 3. Concepts in fractals and percolation. (**A** and **B**) Site lattice percolation for $p < p_c$ (A) and $p > p_c$ (B). White squares are "occupied," black squares are "unoccupied," and blue squares are percolating. (**C**) Illustrative example of a lattice made up of Sierpinski gaskets with correlation length ξ, adopted from (*23*). This lattice is fractal over the short range and homogeneous over the long range. (**D**) MD simulation of the $Cu_{46}Zr_{54}$ system at room temperature with full periodic boundaries (Cu, blue; Zr, yellow). (**E**) $Cu_{46}Zr_{54}$ with all atoms removed, except for those belonging to icosahedrons.

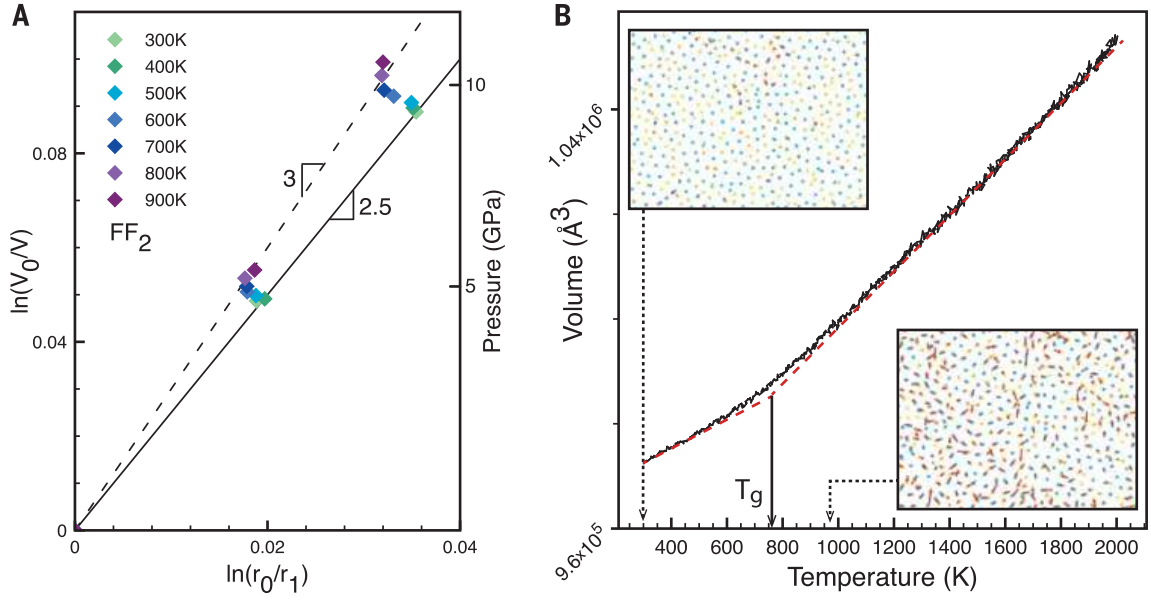


Fig. 4. Simulated properties during supercooling. (**A**) Dimensionality from r_1 during supercooling. (**B**) Volume versus temperature behavior (solid black line), shown with guidelines (red dotted line) and T_g (~763 K, solid black arrow). Inset snapshots show atom vectors (red) generated from reference temperatures ~540 K above the indicated temperatures (dotted black arrows) for a slice 3 Å thick (roughly the nearest-neighbor distance). Dots are atom centers (Cu, blue; Zr, yellow).

previous methods, using multiple data points improved accuracy in measuring the exponent (*10*). Structural information was sensitive to the magnitude of the scattering vector. We found that $D \sim 2.51$ for $\mathbf{q_1}$ (the scattering vector from the first peak position), consistent with previous experiments on other metallic glasses (*10*, *11*). The value of ~2.64 from $\mathbf{q_2}$ measurements was 5% higher than that from $\mathbf{q_1}$ (Fig. 1C). To explore the repeatability of this finding, we analyzed data obtained from a $La_{62}Al_{14}Cu_{11.7}Ag_{2.3}Ni_5Co_5$ metallic glass (*11*). First-peak data for both systems had the same exponent of ~2.5, whereas the $\mathbf{q_2}$ data for the $La_{62}Al_{14}Cu_{11.7}Ag_{2.3}Ni_5Co_5$ had an exponent of nearly 3 (Fig. 1C). This shift is greater than the one observed in the $Cu_{46}Zr_{46}Al_5Be_3$ system, and it supports the observation that a change in the dimensionality arises from probing different extents within the atomic structure in momentum space. Extracting structural information from momentum space measurements is difficult, because the information is spread out. Real-space radial distribution functions (RDFs) are needed, where peak positions correspond directly to atomic separations. Background noise and the limited range of $\mathbf{q}$ restrict the accuracy of Fourier transforms applied to experimental RDFs. Atomistic simulations allow for this type of investigation.

Molecular dynamics (MD) simulations can replicate the glass structure, but the simulation time scales (picoseconds) are many orders of magnitude shorter than in the experiments. We ensured that the system had sufficient time to relax at each pressure increment to address this issue. We held the loading rate constant at 50 GPa/ns (5×10^{19} Pa/s), and we allowed the system to relax for ~0.1 ns to reach thermodynamic equilibrium at each pressure interval. Higher quench rates in simulations may produce less relaxed glasses, although their structures often closely match those produced in experiments (*17*–*19*). The differences in compression rates result in quantitative discrepancies, but the qualitative and phenomenological aspects of the simulations should represent a realistic physical system. We generated $Cu_{46}Zr_{54}$ RDFs by using two embedded-atom-method force fields, described by Cheng *et al.* (FF_1) (*18*) and Mendelev *et al.* (FF_2) (*19*). The neighbor separation–volume relationship for RDF peaks r_1 and r_2 indicated a D of ~2.54, similar to the experimental result, but it transitioned to ~3 between r_2 and r_3 (Fig. 2A). We also simulated $Ni_{80}Al_{20}$, which exhibited a similar crossover between r_1 and r_2 (Fig. 2B) (*20*).

The percolation cluster (*21*) is probably the most relevant fractal model to describe the structure of metallic glasses. The cluster represents a disordered system with fractal dimension $D \sim 2.52$ and appears across many physical systems (*22*, *23*). Percolation models incorporate the probability of occupied (p) and empty ($1 - p$) sites. At low p values, the system is not fully connected (for example, as with an electrical insulator) (Fig. 3A). The percolation threshold (p_c) is reached when a percolating network forms, allowing incipient conduction. Systems characterized by large p have many conduction paths (Fig. 3B). What sets the percolation model apart is the existence of a correlation length, ξ, which characterizes the size of the finite clusters at concentrations below and above p_c. The correlation length is defined as the average distance that spans two sites within the same cluster and has units equal to the size of the smallest constituent unit in the model. At $p < p_c$ and $p > p_c$, ξ is finite, and the system is only fractal at length scales shorter than the correlation length. On length scales longer than ξ, the structure is homogeneous. This property of percolation clusters may help reconcile the notion that fractals need not exhibit self-similarity across all length scales (Fig. 3C) (*23*).

We used a continuum percolation model, where p is analogous to the atomic packing fraction (φ), and the percolation threshold is analogous to a critical volume fraction (φ_c), such that $\varphi_c = \varphi p_c$ (*23*, *24*). The correlation length is

$$\xi \propto |\varphi - \varphi_c|^{-v}$$

for continuum percolation, where $v = 0.8764$ (*25*).

We estimated φ_{CuZr} to be 0.717 to 0.728 using the chemical compositions and the atomic radii of the simulated glasses (*29*). A reasonable model for the packing of a binary metallic glass involves continuum packing of hard spheres, with a p_c of ~0.310 (*26*). We obtained a φ_c of 0.257 by averaging the hard-sphere value ($\varphi_c \sim 0.224 = p_c\varphi_{CuZr}$) with an overlapping sphere value ($\varphi c \sim 0.2896$) (*27*), because atoms in metallic glass are not ideally rigid (*28*). The correlation length was ~2 for $Cu_{46}Zr_{54}$ ($\xi_{CuZr} \sim 1.93$ to 1.98), suggesting that the information in the first and second peaks pertains mostly to the angstrom-sized fractal clusters, whereas information in the third peak pertains to the homogeneous bulk. This result is consistent with our observations of a crossover in dimensionality between r_2 and r_3 (Fig. 2A), and it provides evidence for the presence of percolation structure in metallic glasses. The short-range considerations for high local densities favor the formation of Cu-centered clusters, giving rise to a large number of Cu-centered icosahedra in lieu of the close-packed structures in native Cu and Zr (Fig. 3D) (*16*). The atoms with local icosahedral order form a percolating network (Fig. 3E).

Equation 1 suggests that higher packing fractions bring about shorter correlation lengths. We estimated that the $Ni_{80}Al_{20}$ has a high packing fraction, φ_{NiAl}, of ~0.793, although this could be an overestimation due to the covalent nature of the Al bonding. This estimate gives $\xi_{NiAl} \sim 1.73$, which is much less than 2. The result shows a crossover in dimensionality from ~2.54 to ~3 that occurs between r_1 and r_2 (Fig. 2B). We were also able to induce a shift in the $Cu_{46}Zr_{54}$ crossover from between r_2 and r_3 to between r_1 and r_2 at a pressure of >15 GPa by increasing the packing fraction and bringing ξ_{CuZr} below ~1.7 (fig. S1) (*29*). Some of the atoms in amorphous materials undergo local nonaffine displacements, even in response to purely hydrostatic loads. The fraction of such nonaffine atoms is low (~21.7%), and they do not appear to have any effect on the scaling behavior and crossover (figs. S2 and S3) (*29*).

We related the current model to the glass transition by examining the dimensionality as a function of temperature. We did not observe fractal behavior of $Cu_{46}Zr_{54}$ until 400 K, well below the glass transition temperature (T_g) of 763 K (Fig. 4A). The dimensionality gradually decreased from ~3 to ~2.54 over this temperature range as the temperature decreased. This behavior suggests an intermediary process such as jamming (*9*, *30*, *31*), where the percolating cluster begins to jam at the glass transition. Complete jamming occurs at lower temperatures, along with the emergence of fractal properties, correlating with a loss of ergodicity and consistent with the characteristic kink in the volume-temperature curve during supercooling (Fig. 4B). Despite structural similarities, liquids are amenable to rearrangements in local atomic configurations, whereas in rigid solids, these configurations are preserved. Pressure elicits a mostly nonaffine response from the liquid and a comparatively affine response from the glass. Applied hydrostatic forces inevitably alter the structure and induce structural relaxation in a liquid, which is unavailable in a glass. This difference is probably the reason for the emergence of fractal properties below T_g in a glass and the lack thereof above T_g in a liquid. Metallic liquids possess packing fractions in excess of our estimated percolation threshold, which implies that their atomic structures are also percolating clusters that have not yet frozen or jammed.

A fractal model might be useful in explaining the dynamics of metallic glasses, as concepts from percolation have been applied successfully to other glass formers (*32*). The dynamic heterogeneities that emerge in supercool liquids may be related to the spatial distribution of nonpercolating clusters. Estimating the average number of particles in these clusters using $\sim N_{avg} \approx \xi^3$, where $\xi \sim 2$, we got a value (~8) that is close to experimentally observed values in colloidal glasses (~3 to 7) (*33*). From the perspective of packing, percolation, and jamming, a correlation between density and T_g (*34*) is intuitive. If metallic glasses are created from the jamming of a percolating cluster, then glass formation is simplified: Liquid metal only needs to reach the jamming packing fraction, φ_j, before nucleation occurs. This could be accomplished by a combination of hydrostatic pressure and fast cooling rates. The strong correlation of metallic-glass yield strength with T_g implies that collective atomic motions dictate both yielding and glass formation (*35*). Because denser metallic glasses tend to be better glass formers with higher T_g (*34*), the strength enhancement observed in glasses with higher T_g may originate from the size of the percolating clusters, which increases with packing fraction. Higher packing leads to larger jammed clusters, which present more substantial barriers to the initiation of collective atomic motions that lead to catastrophic shear banding. The movement of

finite nonpercolating clusters may also be related to shear transformation zones, which are collective rearrangements of atoms during the deformation of metallic glasses (*36*). This concept is supported by the observation that typical zone sizes (~10 to 20 atoms) (*37*–*39*) are close to cluster sizes (~8 atoms). The continuum percolation model illustrates how structure and rigidity may organize in the absence of ordering; atoms percolate in the liquid, and the percolating cluster "freezes" (or jams) into a glass.

REFERENCES AND NOTES

1. C. A. Angell, *Science* **267**, 1924–1935 (1995).
2. P. G. Debenedetti, F. H. Stillinger, *Nature* **410**, 259–267 (2001).
3. W. Klement, R. H. Willens, P. Duwez, *Nature* **187**, 869–870 (1960).
4. M. F. Ashby, A. L. Greer, *Scr. Mater.* **54**, 321–326 (2006).
5. D. B. Miracle, *Nat. Mater.* **3**, 697–702 (2004).
6. H. W. Sheng, W. K. Luo, F. M. Alamgir, J. M. Bai, E. Ma, *Nature* **439**, 419–425 (2006).
7. A. Hirata *et al*., *Nat. Mater.* **10**, 28–33 (2011).
8. J. Sietsma, B. J. Thijsse, *J. Non-Cryst. Solids* **135**, 146–154 (1991).
9. A. S. Keys, A. R. Abate, S. C. Glotzer, D. J. Durian, *Nat. Phys.* **3**, 260–264 (2007).
10. D. Ma, A. D. Stoica, X. L. Wang, *Nat. Mater.* **8**, 30–34 (2009).
11. Q. Zeng *et al*., *Phys. Rev. Lett.* **112**, 185502 (2014).
12. B. B. Mandelbrot, *The Fractal Geometry of Nature (Updated and Augmented Edition)* (W.H. Freeman, New York, 1983).
13. T. Tallinen, J. A. Åström, J. Timonen, *Nat. Mater.* **8**, 25–29 (2009).
14. X. Ma, M. R. Zachariah, C. D. Zangmeister, *Nano Lett.* **12**, 486–489 (2012).
15. D. B. Miracle, *Acta Mater.* **61**, 3157–3171 (2013).
16. Y. Q. Cheng, E. Ma, *Prog. Mater. Sci.* **56**, 379–473 (2011).
17. G. Duan *et al*., *Phys. Rev. B* **71**, 224208 (2005).
18. Y. Q. Cheng, E. Ma, H. W. Sheng, *Phys. Rev. Lett.* **102**, 245501 (2009).
19. M. I. Mendelev *et al*., *Philos. Mag.* **89**, 967–987 (2009).
20. G. P. Purja Pun, Y. Mishin, *Philos. Mag.* **89**, 3245–3267 (2009).
21. V. K. S. Shante, S. Kirkpatrick, *Adv. Phys.* **20**, 325–357 (1971).
22. M. Sahimi, *Applications of Percolation Theory* (Taylor & Francis, Bristol, PA, 1994).
23. A. Bunde, S. Havlin, *Fractals and Disordered Systems* (Springer-Verlag, New York, 1991).
24. H. Scher, R. Zallen, *J. Chem. Phys.* **53**, 3759 (1970).
25. J. Wang, Z. Zhou, W. Zhang, T. M. Garoni, Y. Deng, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **87**, 052107 (2013).
26. M. J. Powell, *Phys. Rev. B* **20**, 4194–4198 (1979).
27. C. D. Lorenz, R. M. Ziff, *J. Chem. Phys.* **114**, 3659 (2001).
28. A. S. Clarke, J. D. Wiley, *Phys. Rev. B Condens. Matter* **35**, 7350–7356 (1987).
29. Materials and methods are available as supplementary materials on *Science* Online.
30. A. J. Liu, S. R. Nagel, *Nature* **396**, 21–22 (1998).
31. V. Trappe, V. Prasad, L. Cipelletti, P. N. Segre, D. A. Weitz, *Nature* **411**, 772–775 (2001).
32. R. Orbach, *Science* **231**, 814–819 (1986).
33. E. R. Weeks, J. C. Crocker, A. C. Levitt, A. Schofield, D. A. Weitz, *Science* **287**, 627–631 (2000).
34. Y. Li, Q. Guo, J. A. Kalb, C. V. Thompson, *Science* **322**, 1816–1819 (2008).
35. B. Yang, C. T. Liu, T. G. Nieh, *Appl. Phys. Lett.* **88**, 221911 (2006).
36. M. L. Falk, J. S. Langer, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **57**, 7192–7205 (1998).
37. M. L. Falk, *Phys. Rev. B* **60**, 7062–7070 (1999).
38. A. C. Lund, C. A. Schuh, *Acta Mater.* **51**, 5399–5411 (2003).
39. F. Delogu, *Phys. Rev. Lett.* **100**, 255901 (2008).

ACKNOWLEDGMENTS

Diffraction data and simulated RDFs are available as supplementary materials. The authors thank D. C. Hofmann for providing the $Cu_{46}Zr_{46}Al_5Be_3$ wires and Y. Lin for her aid in sample loading. The authors acknowledge the financial support of the U.S. Department of Energy Office of Basic Energy Sciences (DOE-BES) and NASA's Space Technology Research Grants Program (Early Career Faculty grants to J.R.G.). W.L.M. and C.Y.S. acknowledge support from NSF grant EAR-1055454. Q.Z. acknowledges support from DOE-BES (grant DE-FG02-99ER45775) and the National Natural Science Foundation of China (grant U1530402). Portions of this work were performed at the High Pressure Collaborative Access Team (HPCAT) of the Advanced Photon Source (APS), Argonne National Laboratory. HPCAT operations are supported by DOE's National Nuclear Security Administration (NNSA) under award no. DE-NA0001974 and by DOE-BES under award no. DE-FG02-99ER45775, with partial instrumentation funding by NSF grant MRI-1126249. APS is supported by DOE-BES under contract no. DE-AC02-06CH11357. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for DOE by Stanford University. Some computations were carried out on the Shared Heterogeneous Cluster computers (Caltech Center for Advanced Computing Research) provided by the NNSA Predictive Science Academic Alliance Program at Caltech (grant DE-FC52-08NA28613) and on the NSF Center for Science and Engineering of Materials computer cluster (grant DMR-0520565). Q.A. and W.A.G. received support from the Defense Advanced Research Projects Agency–Army Research Office (grant W31P4Q-13-1-0010) and NSF (grant DMR-1436985). This material is based on work supported by an NSF Graduate Research Fellowship (grant DGE-1144469). Any opinions, findings, and conclusions or recommendations expressed in the material are those of the authors and do not necessarily reflect the views of NSF.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1306/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
References (*40*–*42*)
Databases S1 to S4

14 March 2015; accepted 31 July 2015
10.1126/science.aab1233

APPLIED OPTICS

An ultrathin invisibility skin cloak for visible light

Xingjie Ni,[1]* Zi Jing Wong,[1]* Michael Mrejen,[1] Yuan Wang,[1,2] Xiang Zhang[1,2,3]†

Metamaterial-based optical cloaks have thus far used volumetric distribution of the material properties to gradually bend light and thereby obscure the cloaked region. Hence, they are bulky and hard to scale up and, more critically, typical carpet cloaks introduce unnecessary phase shifts in the reflected light, making the cloaks detectable. Here, we demonstrate experimentally an ultrathin invisibility skin cloak wrapped over an object. This skin cloak conceals a three-dimensional arbitrarily shaped object by complete restoration of the phase of the reflected light at 730-nanometer wavelength. The skin cloak comprises a metasurface with distributed phase shifts rerouting light and rendering the object invisible. In contrast to bulky cloaks with volumetric index variation, our device is only 80 nanometer (about one-ninth of the wavelength) thick and potentially scalable for hiding macroscopic objects.

A cloak is a device that can render objects invisible to incoming waves. Transformation optics and metamaterials provide powerful tools to build cloaking devices. Different schemes relying either on coordinate transformation (*1*–*3*) or scattering cancellation (*4*–*6*) have been studied. Although the concept was first proposed for electromagnetic waves, soon it was extended to acoustic waves (*7*), heat flows (*8*–*10*), elastic or seismic waves (*11*–*13*), and even the matter waves (*14*, *15*). A quasi–conformal mapping technique (*16*) was used to design a so-called carpet cloak that conceals an object by restoring the wavefront as if it were reflected from a flat surface. This technique relaxes the requirements of hard-to-achieve material properties and anisotropy as in the case of the original cloak, thereby making it easier to design and fabricate. Such invisibility carpet cloaks were demonstrated experimentally from microwave (*17*) to optical frequencies (*18*–*23*). Nevertheless, there are still substantial limitations in current optical cloak designs that apply the quasi–conformal mapping technique. Realization requires refractive index modulation over a large volume to avoid extremely high or low index, leading to a bulky cloak. In addition, sophisticated three-dimensional (3D) fabrication with very high spatial resolution is necessary. Therefore, it is challenging to scale up this design to macroscopic sizes. In addition, the varying index has to be less than that of the environment in certain regions (Fig. 1C), making it difficult to create a cloak that works in air. As a result, the cloak is usually embedded in a dielectric prism of higher index which, however, introduces an additional phase in the reflected light and makes the optical cloak itself visible by phase-sensitive detection.

Recent development of metasurfaces pointed out a way to manipulate the phase of a propagating wave directly. The metasurface is an optically thin layer consisting of subwavelength-sized elements that locally tailor the electromagnetic response at the nanoscale accompanied by dramatic light confinement (*24*–*30*). Metasurfaces have enabled a variety of unique phenomena and applications that cannot be achieved conventionally (*31*, *32*)—for example, negative-angle refraction in a broad

[1]NSF Nanoscale Science and Engineering Center (NSEC), University of California, Berkeley, CA 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Department of Physics, King Abdulaziz University, Jeddah 21589, Saudi Arabia.
*These authors contributed equally to this work. **†Corresponding author. E-mail: xiang@berkeley.edu**

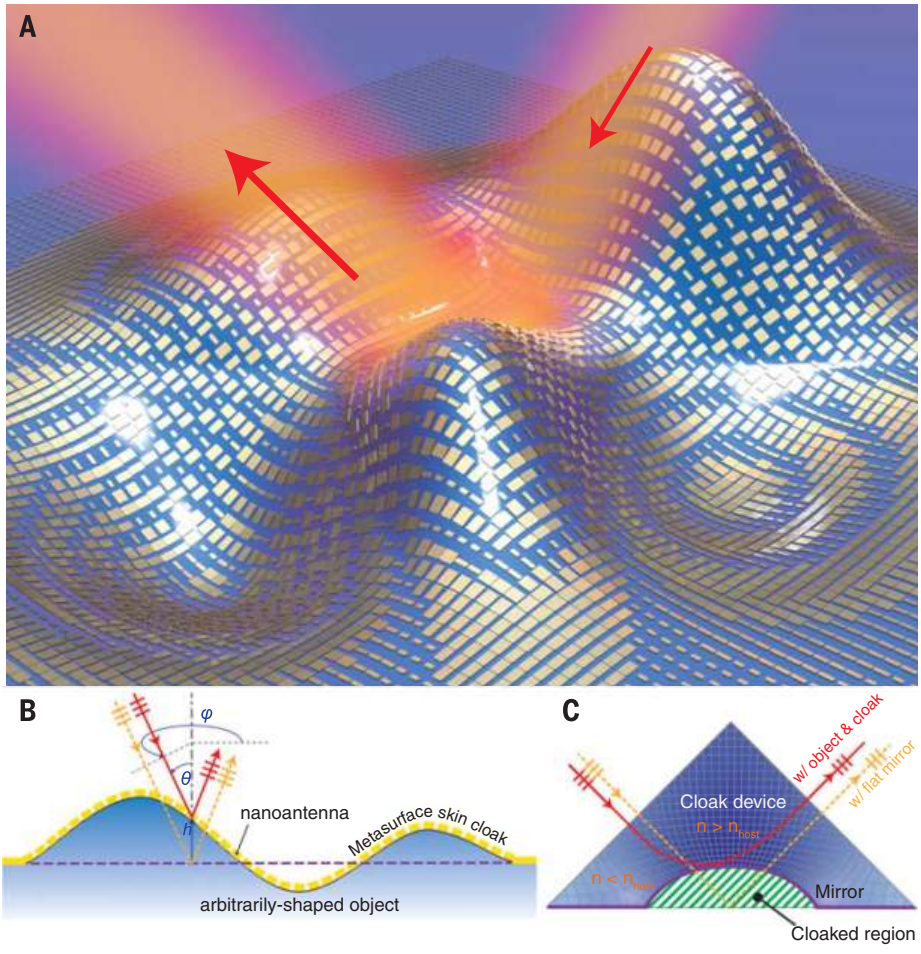


Fig. 1. Schematic view and working principle of a metasurface skin cloak. (**A**) A 3D illustration of a metasurface skin cloak. The skin cloak is an ultrathin layer of nanoantennas (gold blocks) covering the arbitrarily shaped object. (**B**) Schematic view of a cross section of the metasurface skin cloak. (**C**) Schematic illustration of a conventional carpet cloak with spatially varying refractive index (n), designed with an optical quasi–conformal mapping technique. The blue shading qualitatively indicates the local refractive index (the darker color indicates higher index) and the grid lines indicate the deformed optical space. For both devices, the light is incident from the left and exits to the right as represented by the arrows. The devices recover the wavefront (indicated as triple short lines) of the exiting light, such that the object is hidden and looks like a flat mirror, judging from the scattering pattern of the exiting light. However, a conventional cloak introduces additional phase retardation due to the light propagation inside its host material, which renders it detectable to a phase-sensitive measuring device. In contrast, the metasurface skin cloak fully restores both the wavefront and the phase of the scattered light and thus the object is perfectly hidden. On the metasurface skin cloak, a phase shift provided by each nanoantenna realigns the wavefront. At an arbitrary point on the object with height h to the flat reference plane, the nanoantenna should recover the phase of the scattered light. Light (red solid lines) incident at an oblique angle (θ, ϕ) at that point should scatter as if it were reflected by the reference plane (orange dashed lines). The nanoantenna at that point should provide a phase shift $\Delta\Phi = -2k_0 h\cos\theta+\pi$, which compensates the phase difference between the solid and dashed lines. Note that the height is negative when the surface is below the reference plane.

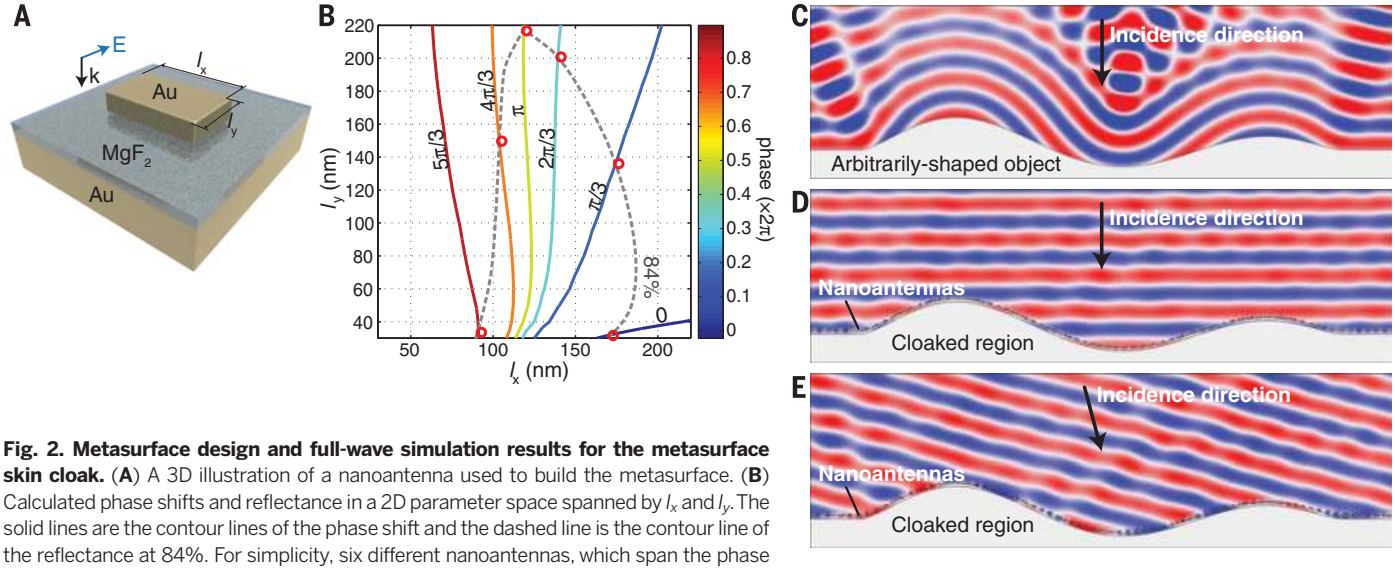


Fig. 2. Metasurface design and full-wave simulation results for the metasurface skin cloak. (**A**) A 3D illustration of a nanoantenna used to build the metasurface. (**B**) Calculated phase shifts and reflectance in a 2D parameter space spanned by l_x and l_y. The solid lines are the contour lines of the phase shift and the dashed line is the contour line of the reflectance at 84%. For simplicity, six different nanoantennas, which span the phase from 0 to 2π with an interval of π/3, are chosen as the building blocks to construct the metasurface, as shown by the six red circles intersecting the phase and the 84%-reflectance contour lines. (**C** to **E**) Full-wave simulated electrical field distribution (shown for a cross section) based on an actual design for (C) a scattering object without a metasurface skin cloak at normal incidence, (D) the same object with a metasurface skin cloak at normal incidence, and (E) the same object with the same metasurface skin cloak as in (D) but at 15° oblique incidence. Only the reflected field is plotted for clarity. The wavelength of the incident light is 730 nm in the simulation. The reflected light is almost completely recovered by the skin cloak as if there were no scattering object for both normal and oblique incidences.

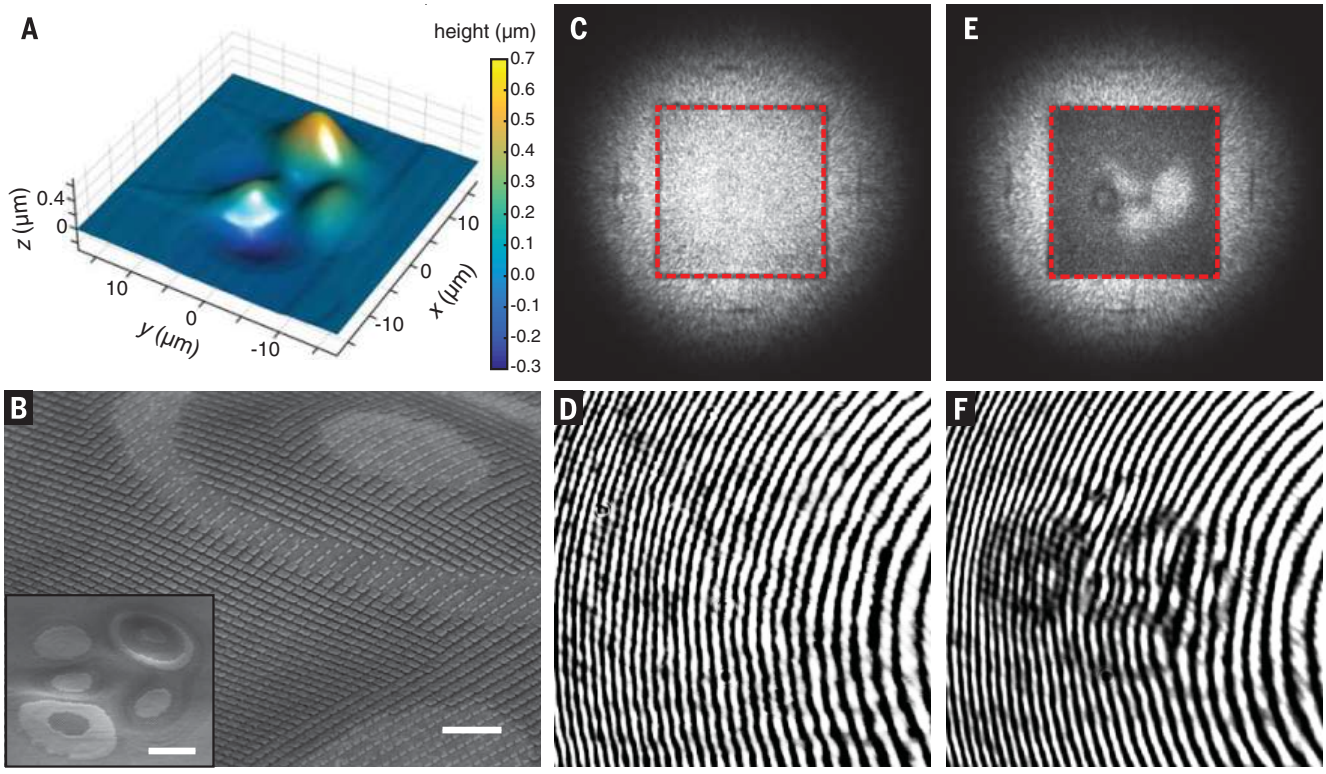


Fig. 3. A metasurface invisibility skin cloak for a 3D arbitrarily shaped object. (**A**) An AFM image of a 3D arbitrarily shaped object that includes multiple bumps and dents. (**B**) A SEM images of the object onto which a metasurface skin cloak has been fabricated. The scale bar is 1 μm. The inset shows an enlarged image of the entire object, with a scale bar of 5 μm. (**C** and **E**) Optical widefield reflection images obtained with a 0.3 numerical aperture objective lens with 730-nm-wavelength laser illumination. The sample region is indicated by the red dashed boxes. (C) The obtained reflection image when the cloak is on and (E) when the cloak is off, taken at the same position. The noise in the images arises from the laser speckle. (**D** and **F**) The respective interference images when the cloak is on and when it is off. When the cloak is off, the interference fringes are distorted on the 3D object, which indicates the height difference on the surface. When the cloak is on, the interference fringes smoothen again, matching with that of the flat surface outside the object region, thus proving that both the wavefront and the phase are well restored without any distortion.

wavelength range (*24*, *25*), unidirectional surface wave coupling (*26*, *33*), planar optical lenses and waveplates (*27*, *34*–*37*), ultrathin high-resolution holograms (*38*–*40*), and enhancement of nonlinear optical responses (*41*).

Here, we experimentally demonstrate an ultrathin invisibility skin cloak (Fig. 1A) at ~730 nm that overcomes the limitations of a bulky cloak. We created a metasurface tightly wrapped over an object to render it free from optical detection. In contrast to the previous bulky design using continuous refractive index distribution over a volumetric space (*16*), the ultrathin layer of this skin cloak reroutes the light and restores the wavefront scattered from the object by compensating the phase difference using phase-shifting resonant elements on the cloak surface. With the complete wavefront and phase recovery, we are able to conceal, for a specific light polarization, a 3D object of arbitrary shape using this skin cloak. Our device has a thickness of 80 nm (~$\lambda/9$) and, being two-dimensional, has much better scalability prospects. Our experiments show that it successfully hides an arbitrarily shaped 3D object of 36 μm by 36 μm at ~730-nm wavelength, both from direct widefield imaging and phase-sensitive detection.

Our metasurface skin cloak consists of subwavelength-scale nanoantennas, which provide distinct phase shifts locally to the reflected electromagnetic waves. On the basis of this phase control capability, we designed the metasurface so that the phase of the scattered light, at each point on the surface of the cloak, would be the same as that of light reflected from a flat mirror. For an oblique-angle (θ, ϕ) light incident on an arbitrarily shaped 3D object at a height of h with respect to the reference plane (Fig. 1B), the introduced phase shift should compensate the phase difference between the light scattered by the object and that reflected from the reference plane. It is straightforward to calculate the phase difference, which is $\Delta\Phi = -2k_0 h\cos\theta + \pi$, where k_0 is the free space wave number. The additional π term represents the phase jump induced by a reflecting mirror. Therefore, nanoantennas designed with local $\Delta\Phi$ phase shift should realign the scattered wavefront. Moreover, the phase should be completely restored, rendering the object undetectable even for phase-sensitive instruments.

As a demonstration, we used a simple rectangular nanoantenna design (Fig. 2A). The calculated phase shifts and reflectance in a 2D parameter space spanned by nanoantenna dimensions l_x and l_y are mapped out for the selection of designs. Six different antennas with phase shifts covering 0 to 2π while preserving the same reflectance were chosen as the building blocks of the metasurface skin cloak. To make the invisibility cloak fully operable required the reflected intensity to be close to that of a mirror. With adaptation of an ultrathin dielectric spacer layer to create a gap plasmon resonance (*42*, *43*), our antennas were designed to attain an overall reflectivity of 84% at around 730-nm wavelength. The parameters for the nanoantennas used in our experiment are indicated as red circles in Fig. 2B. We built a metasurface using those nanoantennas to compensate the local phase change of the reflection. Full-wave simulation results show that strong scattering and substantial wavefront and phase distortion can occur for an arbitrarily shaped 3D object with a maximum height of about 1 μm and a width of about 10 μm (Fig. 2C). However, by wrapping the object with the metasurface skin cloak, the wavefront and phase are completely restored for both normal (Fig. 2D) and slightly oblique incidences (Fig. 2E). Therefore, the object is perfectly hidden even from phase-sensitive detection.

For experimental realization, we used focused ion beam (FIB) to carve out an arbitrarily shaped 3D object with multiple bumps and dents, as illustrated in Fig. 1A. The height profile was mapped out by an atomic force microscope (AFM), and the metasurface skin cloak was then designed on

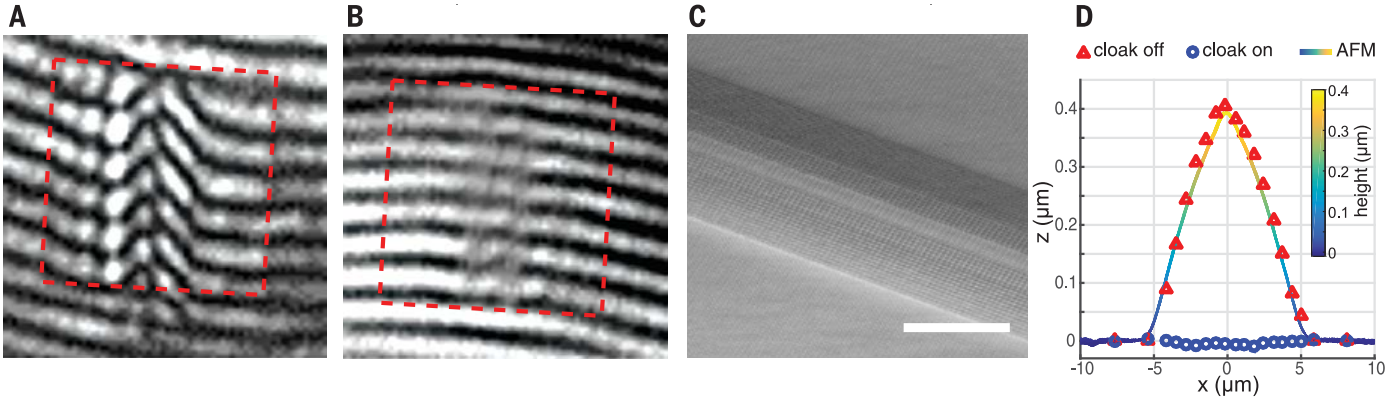


Fig. 4. Quantitative performance evaluation for a metasurface skin cloak on a 3D bump-shaped object. The bump-shaped object has a maximum height of ~400 nm and a width of ~10 μm. (**A** and **B**) Interference fringes obtained when the cloak is respectively off and on. The sample regions are indicated by the red dashed boxes. (**C**) A SEM image of the bump-shaped object with the metasurface skin cloak wrapped over it. The scale bar is 5 μm. (**D**) The extracted height profile from the interference measurement (△ for cloak off and ○ for cloak on) together with that from the AFM measurement (solid line) before fabrication of the nanoantennas.

the basis of this measured height profile (Fig. 3A). The mask was fabricated by means of electron beam lithography with precise alignment. More fabrication details can be found in the supplementary materials. Figure 3B shows the scanning electron microscope (SEM) images of the fabricated metasurface masked on the 3D object. The first test was done by obtaining the widefield reflection image of the device. Because we design our nanoantennas to have the correct phase shift only in the x polarization, we could easily turn the cloak on and off by switching the polarization. With the cloak off, a strong contrast between the cloaked region and the surrounding reflective surface revealed the object (Fig. 3E), but with the correct polarization, the contrast became indistinguishable, rendering the object completely invisible (Fig. 3C). For the phase measurement, a customized Mach-Zehnder interferometer was built (fig. S2) to obtain the reflected phase information from the sample. With the cloak off, there were distortions in the interference fringes indicating nonuniformity in the phase of the reflected light due to the scattering of the object (Fig. 3F). With the cloak on, the distortions disappeared and the fringes were smoothly aligned, indicating that the reflected phase was perfectly uniform over the surface (Fig. 3D). Because we used an objective to image the sample, the interference fringes are rings rather than perfectly straight lines. We also tested an object with an identical profile but without a skin cloak. The interference results (fig. S3B) show fringe distortions similar to those obtained when the cloak is turned off, thereby verifying that the distortion of fringes was purely due to the object scattering and not the metasurface's polarization dependence.

To quantify the performance of a skin cloak, we fabricated and measured a 3D bump-shaped object (Fig. 4). When the cloak was off (Fig. 4A), the interference fringes shifted in the direction perpendicular to the stripes. The extent of phase shift was approximately proportional to the local height of the bump surface. When the cloak was on (Fig. 4B), the interference fringes became smooth and realigned with those from the flat region. No phase jump was observed even at the edges of the skin cloak, proving that there was no additional phase introduced by the cloak. The relative height information was extracted from the extent of fringe dislocation and compared with the AFM height profile measured before the nanoantennas were fabricated (Fig. 4D). When the skin cloak was off, the height from the interference measurement matched well with the AFM results, revealing the true bump profile. Thus, the bump could be precisely mapped out with the use of this phase-sensitive method. When the cloak was on, the extracted height notably dropped to zero over the entire area.

Our skin cloak fully encompasses the advantages associated with metasurfaces. It is extremely thin, because antenna and dielectric spacer thicknesses are 30 and 50 nm, respectively. With a total thickness of only about $^{1}/_{10}$ of the operating wavelengths, it is fully scalable to macroscopic sizes. Theoretically, there is no size limitation, yet in practice, the size can be limited by the macroscopic nanofabrication technology. Nevertheless, meter-scale nanofabrication is now possible—for example, by roll-to-roll nano-imprinting. The cloak can also conceal objects with sharp features like abrupt edges and peaks because the invisibility is attained via local phase adjustments. It is worth pointing out that it is not a strict requirement to put the antennas directly on the object. The metasurface can be on a container of any shape, with or without space between the metasurface and the hidden object. As long as the metasurface is designed correctly, both the container and the objects inside the container will become invisible. In our experiment, we chose to fabricate the antennas right on top of the object to demonstrate the extreme capability of ultrathin 3D cloaking of an arbitrary shape. Our design tolerated incident angles at least within 30°. To allow easy switching between cloak on and cloak off configurations for direct comparison, the metasurface was designed to work only in one polarization. However, by making the nanoantenna shapes symmetric in both x and y directions, it would be possible to implement a polarization-independent invisibility skin cloak. Furthermore, if the nanoantennas could be made adaptive, passively or actively, a deformable skin cloak could potentially conceal any object.

Note added in proof: While this manuscript was in preparation, two theoretical proposals also suggested the use of a phase gradient metasurface to build an invisibility device (*44*, *45*).

REFERENCES AND NOTES

1. J. B. Pendry, D. Schurig, D. R. Smith, *Science* **312**, 1780–1782 (2006).
2. D. Schurig *et al.*, *Science* **314**, 977–980 (2006).
3. W. S. Cai, U. K. Chettiar, A. V. Kildishev, V. M. Shalaev, *Nat. Photonics* **1**, 224–227 (2007).
4. A. Alù, *Phys. Rev. B* **80**, 245115 (2009).
5. P. Y. Fan *et al.*, *Nat. Photonics* **6**, 380–385 (2012).
6. M. Selvanayagam, G. V. Eleftheriades, *Phys Rev. X* **3**, 041011 (2013).
7. L. Sanchis *et al.*, *Phys. Rev. Lett.* **110**, 124301 (2013).
8. T. Han *et al.*, *Phys. Rev. Lett.* **112**, 054302 (2014).
9. H. Xu, X. Shi, F. Gao, H. Sun, B. Zhang, *Phys. Rev. Lett.* **112**, 054301 (2014).
10. R. Schittny, M. Kadic, S. Guenneau, M. Wegener, *Phys. Rev. Lett.* **110**, 195901 (2013).
11. G. W. Milton, M. Briane, J. R. Willis, *New J. Phys.* **8**, 248 (2006).
12. M. Farhat, S. Guenneau, S. Enoch, *Phys. Rev. Lett.* **103**, 024301 (2009).
13. M. Brun, S. Guenneau, A. B. Movchan, *Appl. Phys. Lett.* **94**, 061903 (2009).
14. S. Zhang, D. A. Genov, C. Sun, X. Zhang, *Phys. Rev. Lett.* **100**, 123002 (2008).
15. R. Fleury, A. Alu, *Phys. Rev. B* **87**, 045423 (2013).
16. J. Li, J. B. Pendry, *Phys. Rev. Lett.* **101**, 203901 (2008).
17. H. F. Ma, T. J. Cui, *Nat. Commun.* **1**, 21 (2010).
18. J. Valentine, J. Li, T. Zentgraf, G. Bartal, X. Zhang, *Nat. Mater.* **8**, 568–571 (2009).
19. L. H. Gabrielli, J. Cardenas, C. B. Poitras, M. Lipson, *Nat. Photonics* **3**, 461–463 (2009).
20. T. Ergin, N. Stenger, P. Brenner, J. B. Pendry, M. Wegener, *Science* **328**, 337–339 (2010).
21. M. Gharghi *et al.*, *Nano Lett.* **11**, 2825–2828 (2011).
22. B. Zhang, Y. Luo, X. Liu, G. Barbastathis, *Phys. Rev. Lett.* **106**, 033901 (2011).
23. X. Chen *et al.*, *Nat. Commun.* **2**, 176 (2011).
24. N. Yu *et al.*, *Science* **334**, 333–337 (2011).
25. X. Ni, N. K. Emani, A. V. Kildishev, A. Boltasseva, V. M. Shalaev, *Science* **335**, 427 (2012).

26. S. Sun *et al.*, *Nat. Mater.* **11**, 426–431 (2012).
27. X. Ni, S. Ishii, A. V. Kildishev, V. M. Shalaev, *Light Sci. Appl.* **2**, e72 (2013).
28. L. Huang *et al.*, *Light Sci. Appl.* **2**, e70 (2013).
29. N. Shitrit *et al.*, *Science* **340**, 724–726 (2013).
30. X. Yin, Z. Ye, J. Rho, Y. Wang, X. Zhang, *Science* **339**, 1405–1407 (2013).
31. A. V. Kildishev, A. Boltasseva, V. M. Shalaev, *Science* **339**, 1232009 (2013).
32. N. Yu, F. Capasso, *Nat. Mater.* **13**, 139–150 (2014).
33. J. Lin *et al.*, *Science* **340**, 331–334 (2013).
34. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, *Science* **345**, 298–302 (2014).
35. F. Aieta *et al.*, *Nano Lett.* **12**, 4932–4936 (2012).
36. N. Yu *et al.*, *Nano Lett.* **12**, 6328–6333 (2012).
37. A. Pors, M. G. Nielsen, S. I. Bozhevolnyi, *Opt. Lett.* **38**, 513–515 (2013).
38. L. Huang *et al.*, *Nat. Commun.* **4**, 3808 (2013).
39. X. Ni, A. V. Kildishev, V. M. Shalaev, *Nat. Commun.* **4**, 2807 (2013).
40. G. Zheng *et al.*, *Nat. Nanotechnol.* **10**, 308–312 (2015).
41. J. Lee *et al.*, *Nature* **511**, 65–69 (2014).
42. A. Pors, O. Albrektsen, I. P. Radko, S. I. Bozhevolnyi, *Sci. Rep.* **3**, 2155 (2013).
43. S. Sun *et al.*, *Nano Lett.* **12**, 6223–6229 (2012).
44. N. Mohammadi Estakhri, A. Alù, *IEEE Antennas Wireless Propag. Lett.* **13**, 1775–1778 (2015).
45. L. Y. Hsu, T. Lepetit, B. Kante, *Prog. Electromagnetics Res.* **152**, 33–40 (2015); www.jpier.org/PIER/pier.php?paper=15032005.

ACKNOWLEDGMENTS

The work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05CH11231.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1310/suppl/DC1
Materials and Methods
Figs. S1 to S5
Reference (*46*)

1 July 2015; accepted 17 August 2015
10.1126/science.aac9411

FERROELECTRICS

Emergence of room-temperature ferroelectricity at reduced dimensions

D. Lee,[1] H. Lu,[2] Y. Gu,[3] S.-Y. Choi,[4] S.-D. Li,[5] S. Ryu,[1] T. R. Paudel,[2] K. Song,[6] E. Mikheev,[7] S. Lee,[1] S. Stemmer,[7] D. A. Tenne,[8] S. H. Oh,[6] E. Y. Tsymbal,[2] X. Wu,[5] L.-Q. Chen,[3] A. Gruverman,[2]* C. B. Eom[1]*

The enhancement of the functional properties of materials at reduced dimensions is crucial for continuous advancements in nanoelectronic applications. Here, we report that the scale reduction leads to the emergence of an important functional property, ferroelectricity, challenging the long-standing notion that ferroelectricity is inevitably suppressed at the scale of a few nanometers. A combination of theoretical calculations, electrical measurements, and structural analyses provides evidence of room-temperature ferroelectricity in strain-free epitaxial nanometer-thick films of otherwise nonferroelectric strontium titanate ($SrTiO_3$). We show that electrically induced alignment of naturally existing polar nanoregions is responsible for the appearance of a stable net ferroelectric polarization in these films. This finding can be useful for the development of low-dimensional material systems with enhanced functional properties relevant to emerging nanoelectronic devices.

Low-dimensional ferroelectric structures hold a great potential for scientific and technological endeavors (*1*). Reducing size while retaining ferroelectric properties enables an increase in the storage capacity of nonvolatile ferroelectric memories (*2*), exploration of diverse nanoelectronic functions (*3*–*7*), and discovery of exotic physical phenomena (*8*, *9*). However, maintaining the ferroelectricity in low-dimensional structures, such as ultrathin films, has been hampered by depolarization effects (*10*–*12*), which arise from the uncompensated charges at the interface. The strong scaling effect seems to inevitably suppress ferroelectricity and its functions below a critical dimension (*10*–*13*). A recent theoretical work suggested an intriguing concept for reversibly enhancing ferroelectricity in ultrathin ferroelectric capacitors via the tailoring of chemical bonds at the metal/oxide interface (*14*), but this mechanism has not yet been experimentally confirmed.

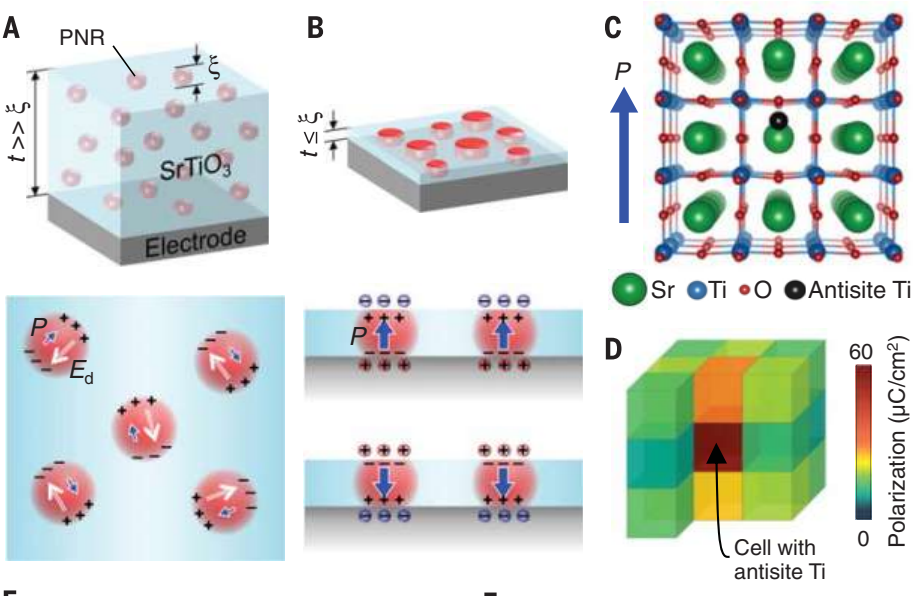
Here, we describe a different mechanism, which enables enhancement of ferroelectricity as the thickness of the system is decreased. In our approach, we use naturally existing polar nanoregions (PNRs)—local nanometer-sized polar clusters—in an archetype dielectric material with perovskite structure: strontium titanate ($SrTiO_3$). PNRs are generally believed to arise from local nanoscale inhomogeneities (such as chemical or structural disorder) (*15*, *16*), which exist in every material (*17*). For example, Sr vacancies are intrinsic point defects in $SrTiO_3$ because of their small formation energy (*18*, *19*), comparable with that of oxygen vacancies, which are likely to act as a natural source of PNRs (Fig. 1A) (*20*, *21*). It has been previously shown (*20*) that relatively thick (tens of nanometers) films of $SrTiO_3$ exhibit relaxor behavior at low temperatures because of the presence of the PNRs. We demonstrate that electrically induced alignment and stabilization of PNRs in nanometer-thick $SrTiO_3$ films results in the emergence of net ferroelectric polarization at room temperature.

Calculations predict that when $SrTiO_3$ is deficient in Sr, antisite Ti defects could instantaneously form and generate local dipole moments by an off-centering displacement (Fig. 1C) (*22*). Our first-principles density-functional theory (DFT) modeling (*23*) shows that the energy gain from this Ti off-centering is as large as ~0.5 eV, originating from structural distortion driven by an ionic radii difference between Ti^{2+} (0.86 Å) and Sr^{2+} (1.44 Å). A local polarization profile around the antisite Ti atom (Fig. 1D) indicates that the off-centered antisite Ti atom induces a large local polarization in its residing unit cell and coherently polarizes the surrounding region. The polarization switchability follows from a calculated energy barrier of ~0.1 eV between the polarization states (Fig. 1E), which is comparable with a double-well potential barrier in conventional ferroelectric materials (*24*). Thus, although pure bulk $SrTiO_3$ is centrosymmetric and nonpolar (*25*), the PNRs of nanometer-characteristic size can naturally form because of the intrinsic Sr deficiency in $SrTiO_3$.

We have previously observed Sr deficiency and associated PNRs even in nominally stoichiometric $SrTiO_3$ bulk single crystals and films (*20*, *21*). These small-sized PNRs, however, do not necessarily generate ferroelectricity. When the film thickness t is much larger than the average PNR size ξ (Fig. 1A), PNRs are isolated in an insulating matrix. The depolarization field E_d in PNRs cannot be effectively screened and thus destabilizes the polarization of PNRs. On the other hand, as t is decreased the PNRs can play a vital role in the emergence of ferroelectricity (Fig. 1B). When t becomes comparable with or smaller than ξ, the electrical boundary conditions for PNRs drastically change as their interfaces come in contact with metallic electrodes and/or become exposed to surface adsorbates. The external charges screen the E_d by compensating for the polarization charge and thus can allow a switchable and stable polarization in PNRs. Such dimensional engineering of polarization stability in PNRs would provide an unconventional way to create and enhance ferroelectricity at reduced dimensions, distinct from methods such as strain (*26*–*28*) and interface (*13*, *14*) engineering.

[1]Department of Materials Science and Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [2]Department of Physics and Astronomy and Nebraska Center for Materials and Nanoscience, University of Nebraska, Lincoln, NE 68588, USA. [3]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA 16802 USA. [4]Department of Materials Modeling and Characterization, Korea Institute of Materials Science, Changwon 642-831, Korea. [5]Department of Physics, Temple University, Philadelphia, PA 19122, USA. [6]Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea. [7]Materials Department, University of California–Santa Barbara, Santa Barbara, CA 93106-5050, USA. [8]Department of Physics, Boise State University, Boise, ID 83725-1570, USA.
***Corresponding author. E-mail: agruverman2@unl.edu (A.G.); eom@engr.wisc.edu (C.B.E.)**

Phase-field simulations were used to model polarization in a representative PNR region with a dimension of 7.5 nm embedded in the $SrTiO_3$ film (*23*). The thickness of the $SrTiO_3$ film in the simulations was varied within the range of 64 unit cells (or 25 nm) to 8 unit cells (3.2 nm) (*23*). We introduced external charges on the top and bottom $SrTiO_3$ interfaces and investigated the stability of the PNR polarization after poling by



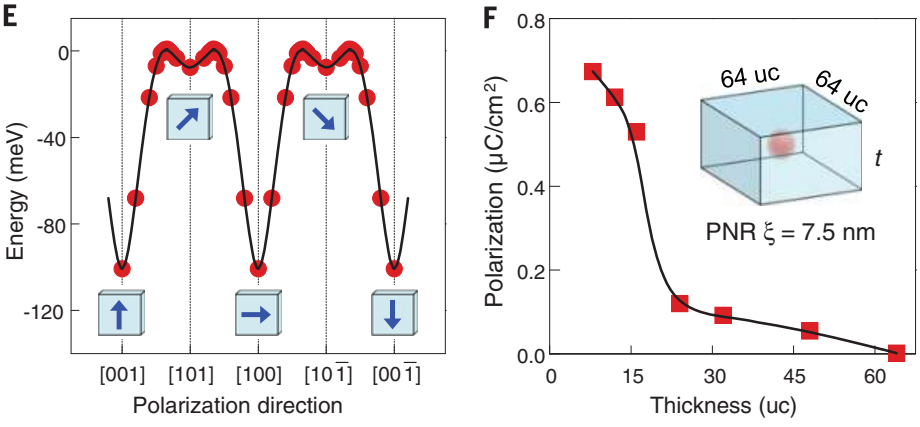


Fig. 1. Theoretical calculations showing the emergence of ferroelectricity in ultrathin films of otherwise nonferroelectric $SrTiO_3$. (**A** and **B**) Dimensional engineering of the polarization (*P*) stability in PNRs. Shown are schematics of PNRs and their *P* in (A) thick $SrTiO_3$ and (B) ultrathin $SrTiO_3$. Blue and white arrows indicate *P* and E_d, respectively. (**C**) Calculated atomic structure near the off-center antisite Ti atom, which induces *P* along the [001] (or its equivalent) direction. (**D**) Profile of local *P* around the antisite Ti atom in a 3 by 3 by 3 supercell obtained from DFT calculations. (**E**) Calculated energy barrier between the polarization states. Polarization switching from [001] to [00$\bar{1}$] direction can be achieved via the metastable polarization states with [101] and [10$\bar{1}$] directions. (**F**) Remnant polarization of the model system, in which a single spherical PNR is embedded in $SrTiO_3$ with a thickness *t*, obtained with phase-field simulations. The lateral dimension of $SrTiO_3$ was fixed as 64 by 64 unit cells (*23*).

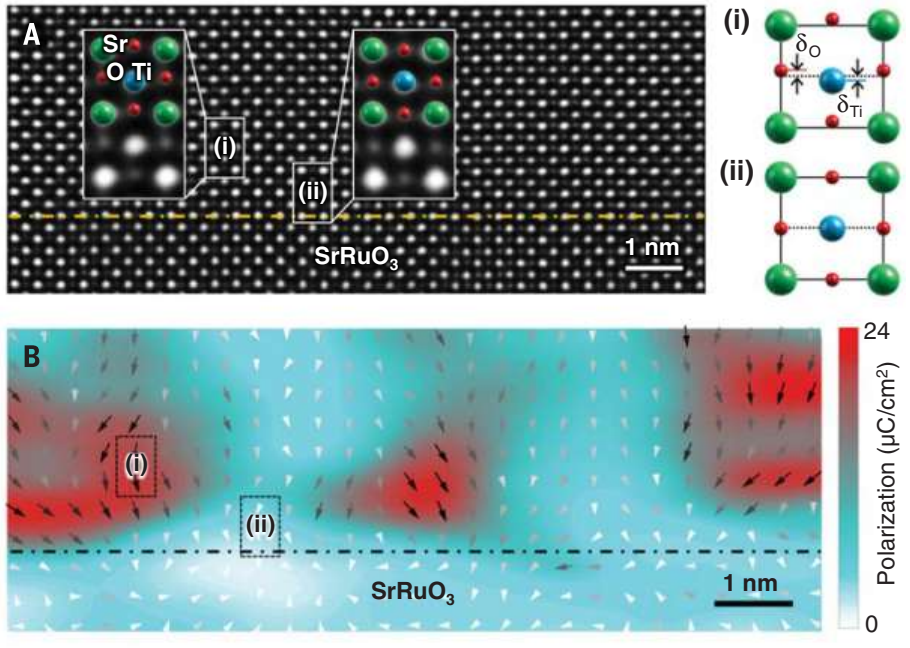


Fig. 2. Atomic-scale imaging of polar nanoregions. (**A**) Filtered ABF-STEM image, including all atomic positions. The (i) and (ii) regions are examples of polar and nonpolar regions, respectively, and are enlarged for clear view (insets). They are schematically drawn on the right; δ_{Ti} and δ_O denote the atomic displacement of Ti and O, respectively. (**B**) Polarization vectors for each unit cell, estimated from atomic displacements in (A). Arrows denote the polarization direction; the stronger the polarization, the darker the arrow color. Strength of polarization is also expressed as a color map, ranging from white (weak) to red (strong).

an external electric field. After the electric field was removed and the system was relaxed to equilibrium, we evaluated the remnant polarization over the whole $SrTiO_3$ film (by the average over the volume). We found that a decrease in the $SrTiO_3$ thickness not only results in a more stable polar state but also greatly enhances the remnant polarization of the system (Fig. 1F). This implies a possible creation and enhancement of ferroelectricity at the reduced scale via the PNR-related mechanism.

We experimentally tested these theoretical predictions in strain-free (001) single-crystal $SrTiO_3$ films (*23*), grown on a (001) $SrTiO_3$ substrate with a single-crystal conductive oxide $SrRuO_3$ bottom electrode (*29*). Film thickness was atomically controlled in the range of 120 to 6 unit cells (fig. S1). The nominally stoichiometric films had a normal unit-cell volume, almost the same as that of the substrate, indicating that the films were nearly free of excessive point defects (fig. S2). Only minute amounts of Sr deficiency (~1 atomic % at most) were present, possibly because of small formation energy (figs. S3 and S4) (*18*, *19*). As mentioned above, Sr deficiency can naturally generate PNRs (Fig. 1, C to E) without compromising crystalline quality of the film, according to the theoretical calculations. Second harmonic generation and Raman spectroscopy measurements provide further evidence of the PNR formation in our $SrTiO_3$ films (*20*, *21*), with a nonpolar-to-polar transition at ~400 K (fig. S3).

To directly visualize the PNRs in our films, we used aberration-corrected scanning transmission electron microscopy (STEM) with high-angle annular-dark-field (HAADF) and annular-bright-field (ABF) techniques (*30*, *31*). We detected the atomic positions of Sr, Ti, and O in the 12-unit-cell-thick film from the ABF-STEM image (Fig. 2A) and the simultaneously obtained HAADF-STEM image (*23*), which allowed us to determine the polar atomic displacements. The magnitude of atomic displacements in polar regions [Fig. 2A, (i)] was as large as ~0.1 Å [comparable with that of conventional ferroelectric $BaTiO_3$ (*24*)] but was negligible in other regions [Fig. 2A, (ii)]. The measured magnitude of atomic displacements is comparable with the calculated average magnitude (~0.06 Å) of atomic displacements in our theoretical model (Fig. 1C). Polarization maps obtained from the Ti and O displacements (*23*) indicate the presence of PNRs with a size of a few nanometers, whereas no clear polarization pattern is observed in the $SrRuO_3$ region (Fig. 2B). PNRs have a downward polarization state in the as-grown film, which is consistent with ferroelectric domain measurements made with piezoresponse force microscopy (PFM) (fig. S6). Following the theoretical modeling and calculations (Fig. 1, B and F), these PNRs may enable ferroelectricity in ultrathin $SrTiO_3$ films, although they cannot generate a macroscopic net polarization in bulk or thick films.

Using a PFM approach, we found that the stable and switchable polarization could indeed be realized at room temperature in ultrathin $SrTiO_3$ films (Fig. 3) (*23*). Bipolar domain patterns, similar to those generated in conventional ferroelectric $BaTiO_3$ films (*5*, *32*), have been created in the $SrTiO_3$ films with a thickness of less than 60 unit cells by scanning the film surface with an electrically biased PFM tip. Bulk-like 120-unit-cell-thick $SrTiO_3$ film did not show any PFM contrast after poling, which is consistent with the notion that PNRs are effective only in the thinnest films. The stability of the written domain patterns was thickness-dependent: Whereas the PFM contrast disappeared within a few minutes in the 60-unit-cell-thick film, the bipolar domain patterns were distinct and stable for several hours or more in thinner films (such as 12-unit-cell-thick film) that exhibited a polarization stability as good as that of ferroelectric 12-unit-cell-thick $BaTiO_3$ films (fig. S8). The PFM result demonstrates that the switchable and stable polarization, which is the signature of ferroelectricity, emerges at room temperature in ultrathin films of otherwise nonferroelectric $SrTiO_3$.

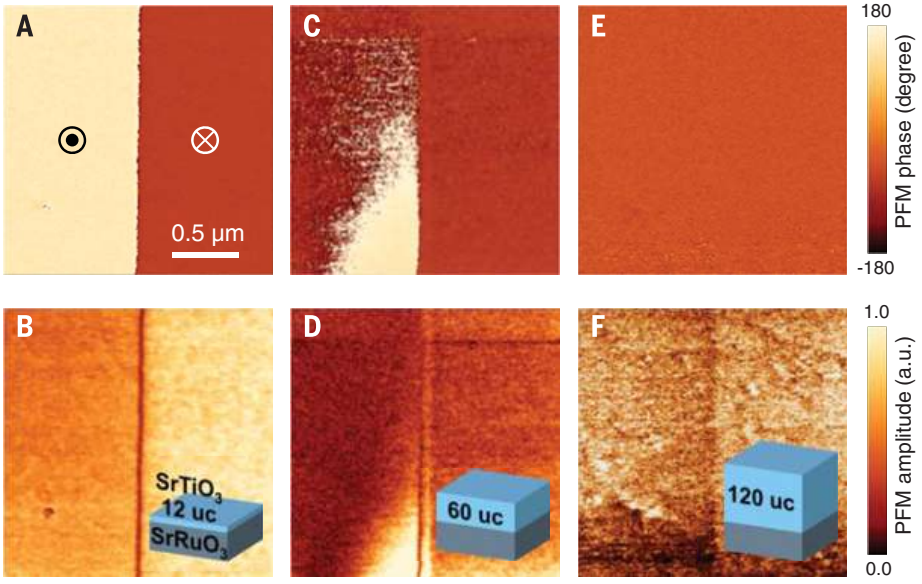


Fig. 3. Visualization of electrically written polarization states in strain-free $SrTiO_3$ films of different thicknesses by means of PFM. (**A** to **D**) Bipolar domain patterns of [(A) and (B)] 12-unit-cell-thick and [(C) and (D)] 60-unit-cell-thick $SrTiO_3$ films. The upward polarization in (C) and (D) spontaneously relaxes within several minutes. The downward polarization in (C) and (D) is characterized by a weaker amplitude PFM signal. The 24- and 6-unit-cell-thick $SrTiO_3$ films also show a bipolar domain pattern with clear contrast (*23*). (**E** and **F**) The thickest $SrTiO_3$ film of 120 unit cells does not exhibit any contrast after domain writing. The PFM amplitude signal is weak, which is consistent with the nonpolar nature of the bulk-like $SrTiO_3$ films.

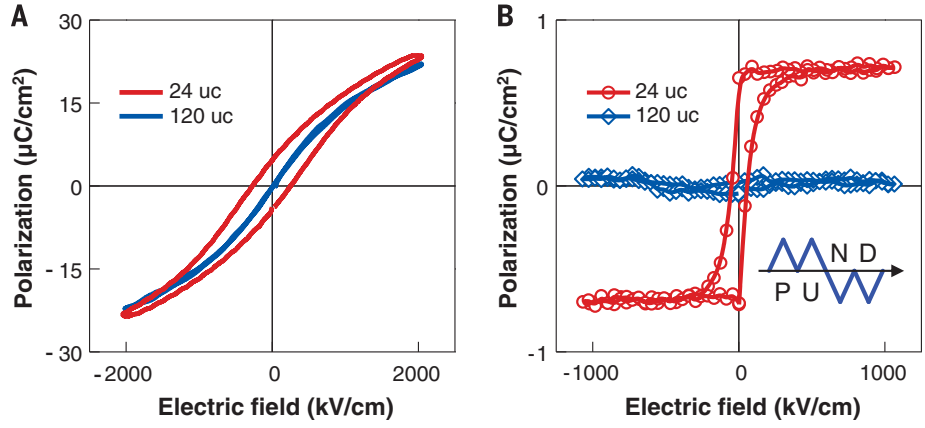


Fig. 4. Polarization hysteresis in ultrathin $SrTiO_3$ films. (**A**) Polarization versus electric field curves measured at room temperature for 24- and 120-unit-cell-thick $SrTiO_3$ films. We performed hysteresis measurements at the frequency of 20 kHz. (**B**) The polarization hysteresis of 24- and 120-unit-cell-thick $SrTiO_3$ films at room temperature, measured by using the double-wave method with a triangular ac electric field of 10 kHz. (Inset) Schematic of applied waveform. We obtained the pure hysteresis component by subtracting the nonhysteresis polarization [up (U) and down (D) runs] from the total [positive (P) and negative (N) runs].

Further evidence for room-temperature ferroelectricity in ultrathin $SrTiO_3$ films was obtained directly from polarization hysteresis (P–E) loops. The 24-unit-cell-thick $SrTiO_3$ films showed a clear polarization hysteresis with nonzero remnant polarization, whereas bulk-like 120-unit-cell-thick $SrTiO_3$ films showed no hysteresis with a paraelectric-like behavior (Fig. 4A). The measured P–E curves included a strong nonlinear dielectric contribution. To eliminate this response and to enable a more accurate measurement, we used the double-wave method (*33*), giving a switched polarization $\Delta P = P^+ - P^-$ of ~1.4 μC/cm^2 for the 24-unit-cell-thick $SrTiO_3$ films but $\Delta P \approx 0$ for the 120-unit-cell-thick films (Fig. 4B). Considering the low density of PNRs in our films, the measured ΔP value seems reasonable, but we believe that it can be increased through interfacial engineering for more efficient charge compensation (*13*, *14*). Our PFM and hysteresis measurements unambiguously confirm room-temperature ferroelectricity in ultrathin $SrTiO_3$ films as well as the enhancement of ferroelectricity at the nanoscale range in agreement with the theoretical predictions (Fig. 1).

This study demonstrates that size reduction does not necessarily lead to the deterioration of ferroelectric properties but in fact could enhance them. Although this mechanism might be limited to the ferroelectrics with a relatively low value of polarization and relaxors, it provides a path toward devices with reduced dimensions in which ferroelectricity is coupled to other functional properties, such as two-dimensional conductivity (*34*), superconductivity (*35*), and magnetism (*36*). In particular, we envision nonvolatile devices with ferroelectric polarization controlling interfacial carrier concentrations. Not limited to $SrTiO_3$, our approach can be applied to other perovskite dielectrics (*37*), in which PNRs are controlled through defect engineering, as well as artificially layered superlattices (*38*, *39*), in which PNRs can influence multiple interfaces.

REFERENCES AND NOTES

1. C. H. Ahn, K. M. Rabe, J.-M. Triscone, *Science* **303**, 488–491 (2004).
2. J. F. Scott, C. A. Paz de Araujo, *Science* **246**, 1400–1405 (1989).
3. M. Y. Zhuravlev, R. F. Sabirianov, S. S. Jaswal, E. Y. Tsymbal, *Phys. Rev. Lett.* **94**, 246802 (2005).
4. V. Garcia *et al.*, *Nature* **460**, 81–84 (2009).
5. A. Gruverman *et al.*, *Nano Lett.* **9**, 3539–3543 (2009).
6. V. Garcia *et al.*, *Science* **327**, 1106–1110 (2010).
7. D. Pantel, S. Goetze, D. Hesse, M. Alexe, *Nat. Mater.* **11**, 289–293 (2012).
8. E. Durgun, P. Ghosez, R. Shaltaf, X. Gonze, J.-Y. Raty, *Phys. Rev. Lett.* **103**, 247601 (2009).
9. J. Mannhart, D. G. Schlom, *Science* **327**, 1607–1611 (2010).
10. W. L. Zhong, Y. G. Wang, P. L. Zhang, B. D. Qu, *Phys. Rev. B Condens. Matter* **50**, 698–703 (1994).
11. J. Junquera, P. Ghosez, *Nature* **422**, 506–509 (2003).
12. C. Lichtensteiger, J.-M. Triscone, J. Junquera, P. Ghosez, *Phys. Rev. Lett.* **94**, 047603 (2005).
13. H. Lu *et al.*, *Adv. Mater.* **24**, 1209–1216 (2012).
14. M. Stengel, D. Vanderbilt, N. A. Spaldin, *Nat. Mater.* **8**, 392–397 (2009).
15. G. Burns, F. H. Dacol, *Phys. Rev. B* **28**, 2527–2530 (1983).
16. L. E. Cross, *Ferroelectrics* **76**, 241–267 (1987).
17. N. W. Ashcroft, N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976).
18. K. R. Udayakumar, A. N. Cormack, *J. Phys. Chem. Solids* **50**, 55–60 (1989).
19. B. Liu *et al.*, *Phys. Chem. Chem. Phys.* **16**, 15590–15596 (2014).
20. H. W. Jang *et al.*, *Phys. Rev. Lett.* **104**, 197601 (2010).
21. D. A. Tenne *et al.*, *Appl. Phys. Lett.* **97**, 142901 (2010).
22. M. Choi, F. Oba, I. Tanaka, *Phys. Rev. Lett.* **103**, 185502 (2009).
23. Materials and methods are available as supplementary materials on *Science* Online.
24. R. E. Cohen, *Nature* **358**, 136–138 (1992).
25. K. A. Müller, H. Burkard, *Phys. Rev. B* **19**, 3593–3602 (1979).
26. J. H. Haeni *et al.*, *Nature* **430**, 758–761 (2004).
27. M. P. Warusawithana *et al.*, *Science* **324**, 367–370 (2009).
28. K. J. Choi *et al.*, *Science* **306**, 1005–1009 (2004).
29. C. B. Eom *et al.*, *Science* **258**, 1766–1769 (1992).
30. K. W. Urban, *Science* **321**, 506–510 (2008).
31. R. Ishikawa *et al.*, *Nat. Mater.* **10**, 278–281 (2011).
32. H. Lu *et al.*, *Science* **336**, 59–61 (2012).
33. J. F. Scott *et al.*, *J. Appl. Phys.* **64**, 787–792 (1988).
34. A. Ohtomo, H. Y. Hwang, *Nature* **427**, 423–426 (2004).
35. N. Reyren *et al.*, *Science* **317**, 1196–1199 (2007).
36. A. Brinkman *et al.*, *Nat. Mater.* **6**, 493–496 (2007).
37. U. Bianchi, J. Dec, W. Kleemann, J. G. Bednorz, *Phys. Rev. B Condens. Matter* **51**, 8737–8746 (1995).
38. D. A. Tenne *et al.*, *Science* **313**, 1614–1616 (2006).
39. A. T. Bollinger *et al.*, *Nature* **472**, 458–460 (2011).

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) under Designing Materials to Revolutionize and Engineer our Future (DMREF) grant DMR-1234096. The research at University of Nebraska–Lincoln was supported by NSF through the Materials Research Science and Engineering Center (MRSEC) under grant DMR-1420645. S.Y.C. acknowledges the support of the Global Frontier Hybrid Interface Materials of the National Research Foundation of Korea funded by the Korea Government (2013M3A6B1078872). The work at Penn State is partially supported by Penn State NSF-MRSEC Center for Nanoscale Science grant DMR-1420620 and by NSF grant DMR-1410714. The work at Temple University was supported as part of the Center for the Computational Design of Functional Layered Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences under award DE-SC0012575 and used resources of the National Energy Research Scientific Computing Center (NERSC), a User Facility supported by the DOE Office of Science. S.H.O. and K.S. acknowledge the support of the Asian Office of Aerospace Research and Development (AOARD) under grant FA2386-15-1-4046, Brain Korea 21 PLUS project for Center for Creative Industrial Materials (grant F14SN02D1707), and the National Research Foundation (NRF) of Korea funded by the Korean Government (grant 2015R1A2A2A01007904). Raman studies at Boise State University have been supported by NSF under grant DMR-1006136. The work at University of California–Santa Barbara was partially supported by the MRSEC Program of the National Science Foundation under award DMR-1121053.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1314/suppl/DC1
Materials and Methods
Figs. S1 to S9
References (*40*–*53*)

20 January 2015; accepted 3 August 2015
10.1126/science.aaa6442

QUANTUM OPTICS

Deterministic generation of multiparticle entanglement by quantum Zeno dynamics

Giovanni Barontini, Leander Hohmann, Florian Haas,* Jérôme Estève, Jakob Reichel†

Multiparticle entangled quantum states, a key resource in quantum-enhanced metrology and computing, are usually generated by coherent operations exclusively. However, unusual forms of quantum dynamics can be obtained when environment coupling is used as part of the state generation. In this work, we used quantum Zeno dynamics (QZD), based on nondestructive measurement with an optical microcavity, to deterministically generate different multiparticle entangled states in an ensemble of 36 qubit atoms in less than 5 microseconds. We characterized the resulting states by performing quantum tomography, yielding a time-resolved account of the entanglement generation. In addition, we studied the dependence of quantum states on measurement strength and quantified the depth of entanglement. Our results show that QZD is a versatile tool for fast and deterministic entanglement generation in quantum engineering applications.

Engineering a desired quantum state—including the increasingly complex entangled states required for quantum computing and quantum simulations—is typically accomplished using only coherent interactions, such as a resonant field driving an atomic transition. Recent developments show that measurement and coupling to environment can also be used as powerful tools for quantum engineering (*1*–*7*). One example of this concept is quantum Zeno dynamics (QZD), which has been theoretically studied for more than a decade (*8*–*10*). QZD combines a coherent interaction with a measurement that is applied simultaneously. In the simplest case, this measurement detects just a single, initially unoccupied state in the large state space in which the system evolves. If the measurement is performed frequently enough, measurement back-action will keep the detected state unoccupied indefinitely; this is the well-known quantum Zeno effect. However, the simple fact of blocking one state also involves a profound modification of the dynamics elsewhere in the state space: This dynamics, called QZD, remains coherent but produces states that

Laboratoire Kastler Brossel, École Normale Supérieure, Université Pierre et Marie Curie Paris 6, CNRS, Collège de France, 24 rue Lhomond, 75005 Paris, France.
*Present address: TWS Partners, 80538 München, Germany.
†Corresponding author. E-mail: jakob.reichel@ens.fr

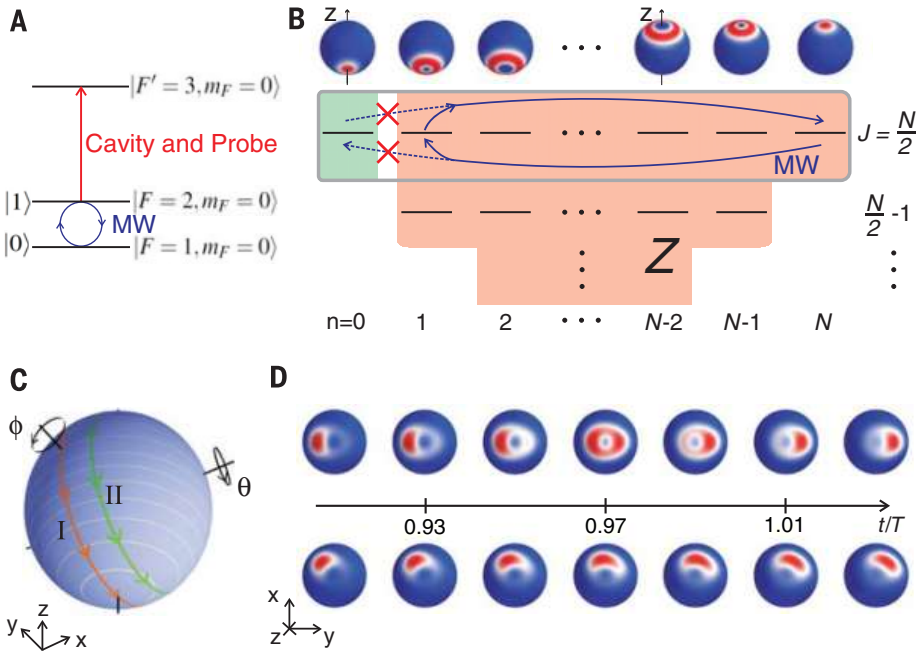


Fig. 1. Quantum Zeno dynamics of atoms coupled to a cavity. (**A**) Relevant level scheme of ^{87}Rb. A resonant 6.8-GHz microwave (MW) allows the application of arbitrary rotations to the atomic qubit. The cavity and probe laser are resonant with the transition $|1\rangle \rightarrow |F'=3, m_{F'}=0\rangle$ (F' and $m_{F'}$ are the hyperfine and magnetic quantum number of the excited state, respectively). The cavity transmits only when the atoms are in the $|0_N\rangle$ state. (**B**) The Hilbert space of the symmetric atomic spin states is spanned by the Dicke states $|n_N\rangle$ (see text). The Husimi-Q distributions of some of these states are also displayed. Starting with the atoms in $|N_N\rangle$, we apply the microwave and simultaneously measure cavity transmission, leading to coherent evolution that is restricted to the subspace Z (orange shaded area). (**C**) Microwave excitations used in the experiment. Plotted are the Bloch sphere trajectories of the mean spin without measurement. Trajectory I drives the mean spin through the south pole, whereas trajectory II avoids the pole. (**D**) Simulated evolution of the Husimi-Q distribution on the Bloch sphere for 36 atoms on trajectory I (upper row) and trajectory II (lower row) under cavity measurement.

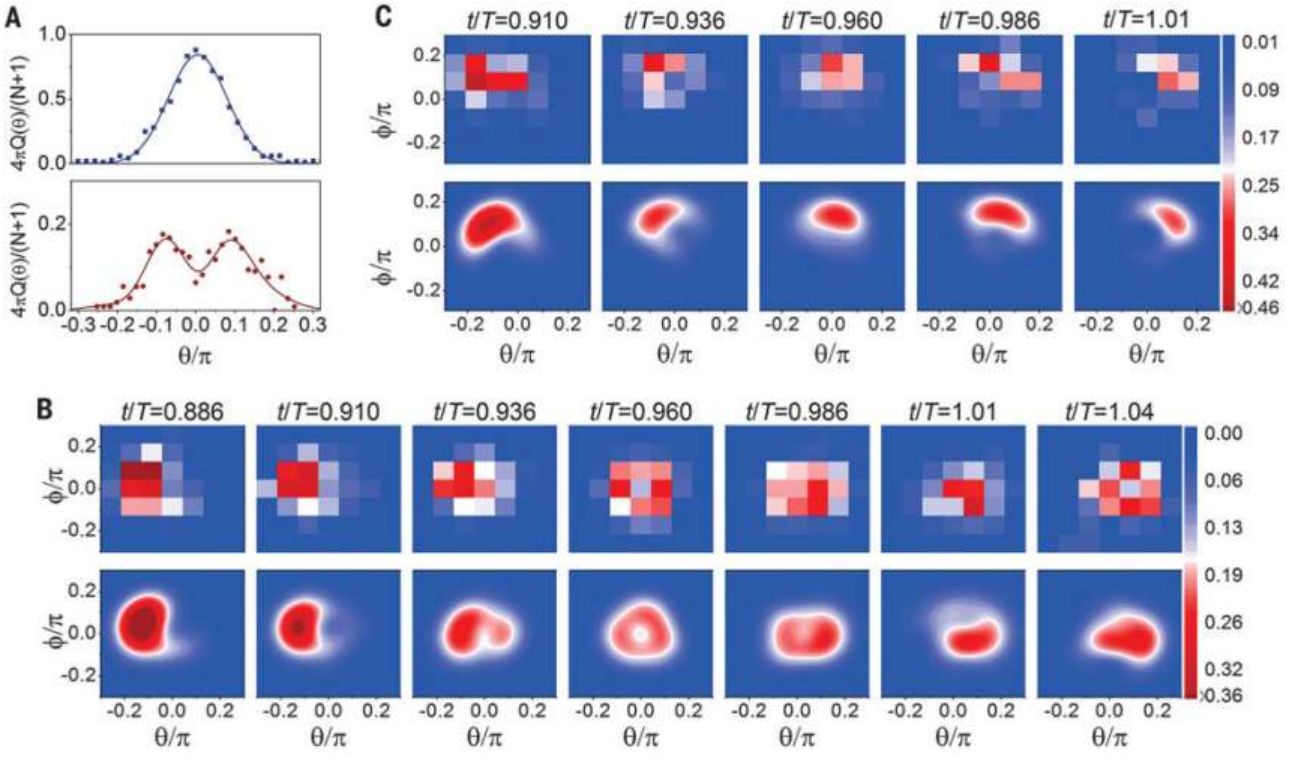


Fig. 2. Tomography of quantum Zeno dynamics. (**A**) Measured high-resolution 1D cuts of the Husimi-Q distributions $Q(\theta, \phi = 0)$ when the mean spin reaches the south pole. The upper panel shows the state after microwave drive along trajectory I without measurement. The Husimi-Q function displays the $\cos(\theta/2)^N$ dependence of the coherent state. The lower panel results from the same microwave drive, but in the presence of measurement. The Husimi-Q function shows a characteristic dip. The lines are maximum likelihood reconstructions (*16*). (**B**) Measured 2D distributions $Q(\theta, \phi)$ at different times for trajectory I in the presence of measurement. The 7-by-7 measurement grid is centered around the south pole. The lower row shows the Husimi-Q distribution corresponding to the reconstructed density matrices. For $t/T = 0.96$, a ring-shaped distribution appears, indicating high overlap with the W state. (**C**) Same as (B), but for trajectory II. The quantum Fisher information normalized to the initial atom number $N = 36$ is (from left to right) $1.51^{+0.39}_{-0.08}, 1.25^{+0.3}_{-0.17}, 1.47^{+0.14}_{-0.42}, 1.19^{+0.25}_{-0.14}, 1.14^{+0.25}_{-0.11}$. The measurement rate is $r_m \simeq 22.5\Omega$ for all panels.

can be completely different from those produced by the same coherent interaction without the measurement (*8*–*10*). QZD may result in quantum states that would be inaccessible in the absence of measurement and that are potentially interesting for quantum engineering. The experimental challenge is to realize the required nondemolition measurement, which must have high efficiency for one state (or a group of states) without affecting the fragile quantum coherences in the rest of the state space. Single-particle QZD has been realized by inducing losses on a specific atomic state (*11*). Other recent experiments with single Rydberg atoms (*12*) and photons in a superconducting microwave cavity (*13*) have used unitary operators to divide the state space and thus obtain an equivalent effect without the use of measurement.

Here we use a cavity-based measurement in the regime of high cooperativity, $C \gg 1$, where $C = g^2/2\kappa\gamma$ (g is the single-atom coupling rate to the cavity field; κ and γ are the atomic and cavity half-linewidths at half maximum, respectively). Compared with other forms of optical detection, the cavity reduces the spontaneous emission rate by orders of magnitude, making this a good approximation of an ideal projective measurement (*14*–*16*). An ensemble of ^{87}Rb atoms is confined in a single antinode of the cavity field, ensuring a near-identical coupling rate $g = 2\pi \times 190$ MHz. Two hyperfine ground-state sublevels serve as qubit states $|0\rangle$ and $|1\rangle$ (Fig. 1A), and a microwave source is used to drive transitions between these states, realizing a unitary operation U_{MW}. Because the cavity and microwave fields both provide near-identical coupling to all of the atoms, the relevant Hilbert space consists of the symmetric states of the collective atomic spin, which is spanned by the Dicke states $|n_N\rangle$ (Fig. 1B), where n is the number of atoms in $|1\rangle$ and N is the total atom number (*16*). In our experiment, the cavity and probe beam are tuned as shown in Fig. 1A, so that the measurement distinguishes the state $|0_N\rangle$ (all atoms in $|0\rangle$), for which the cavity transmits, from the subspace Z of all other states, for which it reflects (*15*). When U_{MW} takes the N-atom state close to the boundary of Z, the dynamics in presence of the cavity transmission measurement strongly differs from U_{MW} applied alone. The states created in the presence of the measurement are highly entangled in general, and their purity depends on the measurement being strong enough and, at the same time, sufficiently nondestructive. For example, starting from the initial state $|N_N\rangle$ (all atoms in $|1\rangle$), driving microwave Rabi oscillations with Rabi frequency Ω (trajectory I in Fig. 1C) would normally produce the state $|0_N\rangle$ after an evolution time $T = \pi/\Omega$. However, if the cavity measurement is applied simultaneously with the microwave drive, the system cannot reach this state. Instead, it evolves through a state that is very close to the W state, $|1_N\rangle = 1/\sqrt{N}(|10\ldots0\rangle + |010\ldots0\rangle + \ldots + |00\ldots1\rangle)$, an entangled state that is robust against particle loss and enables some metrological gain over nonentangled states (*17*, *18*). If the microwave drive follows trajectory II in Fig. 1C, such that U_{MW} would go through a state different than but close to $|0_N\rangle$, then a different entangled state is obtained. Figure 1D shows simulated states of the N atoms for the two trajectories in the presence of measurement, in the vicinity of the π-pulse time. Plotted is the Husimi-Q distribution (*15*) of the states, which for symmetric states is defined as $Q(\theta,\phi) = (N+1)/(4\pi)P_0(\theta,\phi)$, where $P_0(\theta,\phi) = \langle 0_N|R^{\dagger}_{\theta,\phi}\rho R_{\theta,\phi}|0_N\rangle$ is the probability that all atoms are in $|0_N\rangle$ after a rotation $R_{\theta,\phi}$ of the state with density matrix ρ.

In our experiment, we used an ensemble of 36 ± 2 atoms (*16*). To track the dynamics, we took snapshots of the atomic state by stopping the microwave and the measurement at different times t during the evolution. For each t, we measured $Q(\theta,\phi)$ (*15*, *16*) of the resulting state. In the absence of measurement, driving the atoms along trajectory I, we obtain the usual Rabi oscillation (Fig. 2A, upper panel). By contrast, in the presence of measurement, we observe that the Husimi-Q distribution is deformed and features the expected dip in the center (lower panel in Fig. 2A) when the state reaches the south pole. A detailed comparison to theory is discussed below. In Fig. 2, panels B and C show the result of full two-dimensional (2D) tomography for different times on trajectories I and II, respectively. In both cases, as the state approaches the boundary of Z, its Q function is deformed such that $Q(0,0)$, which is proportional to the population in $|0_N\rangle$, remains small at all times. Driving the dynamics further, the state recovers its Gaussian character as it leaves the measurement boundary.

From the tomographic measurements, we reconstructed the symmetric part ρ_s of the atomic ensemble density matrix ρ using a maximum likelihood method (*15*, *16*). On trajectory I, the Q function completely encircles the forbidden state for $t = 0.96T$, where $T = 4.65$ μs. At this time, the population ρ_{11} of the W state reaches a maximum of 0.37 ± 0.04, whereas $\rho_{00} = 0.17 \pm 0.03$. Knowing ρ_{11} and ρ_{00}, we can use the entanglement-depth criterion derived in (*15*): Any N-particle ρ that has a given combination $\{\rho_{00}, \rho_{11}\}$, when decomposed into smaller density matrices $\rho_1 \otimes \ldots \otimes \rho_M$, must contain at least one ρ_i with $\dim\rho_i \geq k$, where k depends on the values $\{\rho_{00}, \rho_{11}\}$. This criterion gives the minimum number of atoms that are demonstrably entangled with one another but makes no statement about the strength of this entanglement, which might be weak. Such a criterion of k-particle entanglement, which does not involve the exact knowledge of ρ_i, was first derived for spin-squeezed

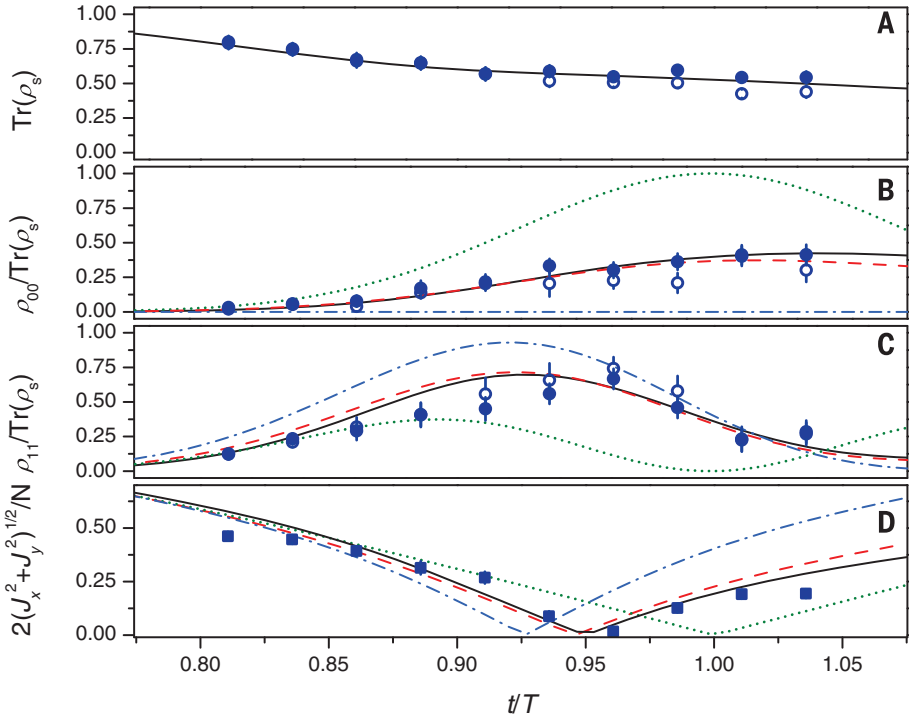


Fig. 3. Evolution of the Dicke state populations during QZD along trajectory I. The measured populations (solid symbols) are deduced from the data in Fig. 2B. (**A**) Population inside the symmetric subspace. (**B** and **C**) Relative populations of the Dicke states $|0_N\rangle$ (B) and $|1_N\rangle$ (C). (**D**) Transverse spin length $2/N\sqrt{J_x^2 + J_y^2}$, where J_i is the ith component of the collective atomic spin. The green dotted lines in (B) to (D) show the expected evolution without measurement (Rabi oscillation). The measured data are well described by a model including spontaneous emission (*16*) with no adjustable parameters (black solid lines). The dynamics within the symmetric subspace can also be understood without considering spontaneous emission (red dashed lines). Blue dot-dashed lines are predictions for ideal QZD ($r_m \to \infty$). Open symbols represent measured evolution, excluding runs with nonzero cavity transmission during the QZD. Error bars are 1σ statistical errors of the reconstruction.

states (*19*). Our $k(\rho_{00}, \rho_{11})$, derived in (*15*), efficiently detects entanglement in the vicinity of the W state. Applied to the state for $t/T = 0.96$, $k(\rho_{00}, \rho_{11})$ reveals that this state contains at least 8^{+3}_{-5} entangled particles, in spite of the experimental imperfections discussed below.

For trajectory II, the resulting distributions are compressed in one direction with respect to a coherent state. For such states, our entanglement-depth criterion is not efficient. To assess their nonclassical character, we calculate a lower bound on their quantum Fisher information F_Q (*16*, *18*). For ideal QZD (infinite measurement rate and no loss) on this trajectory, numerical simulations indicate that a maximum $F_Q/N \simeq 3.4$ is reached for $t/T \simeq 0.96$. It has been shown that $F_Q > N$ is a sufficient (but not necessary) condition for entanglement; also, F_Q/N is the maximum achievable reduction in the variance of an estimate of a quantum phase using that state, with respect to the shot noise limit (*18*, *20*). The experimentally realized states in Fig. 2C yield $F_Q/N = \{1.51, 1.25, 1.47, 1.19, 1.14\}$ (*16*), showing that this trajectory also creates states that are entangled and feature some metrological gain with respect to classical states.

The cavity measurement is characterized by an effective rate that is not infinitely high, and this measurement is accompanied by spontaneous emission, which tends to populate undesired states outside the symmetric subspace. Experimentally, the symmetric subspace population $\mathrm{Tr}(\rho_s)$ is readily available from the reconstructed density matrices. Figure 3A shows the decay of $\mathrm{Tr}(\rho_s)$ for trajectory I. A model (*16*) including spontaneous emission (hereafter referred to as the "full model"), with no adjustable parameters, reproduces the data well (solid line). Apart from this decay, the features of the QZD can be understood without accounting for spontaneous emission. In Fig. 3, panels B and C show the measured relative populations $\rho'_{ii} = \rho_{ii}/\mathrm{Tr}(\rho_s)$ (solid symbols), comparing them with ideal QZD (dot-dashed lines) and QZD with a finite measurement rate (dashed lines). ρ'_{00}, which would reach 1 in the absence of measurement, is strongly reduced by the measurement (Fig. 3B), whereas ρ'_{11} is increased (Fig. 3C). We also observe that the turning point of the Rabi oscillation of the collective spin shifts to shorter times (Fig. 3D); this is expected because the measurement reduces the dimension of the Hilbert space. The deviation from the ideal QZD is well described by a simpler model (dashed lines), which takes into account only the finite measurement rate, not the spontaneous emission. The atoms coherently evolve according to the Hamiltonian $H_{MW}/\hbar = \Omega J_x$ (where $\hbar$ is Planck's constant h divided by 2π, and J_x is the x component of the angular momentum operator) and are subject to quantum jumps with a single jump operator $d = \sqrt{r_m}|0_N\rangle\langle 0_N|$ accounting for the effect of the measurement (*16*). The measurement rate is $r_m = 2\Phi\sqrt{T_0}$, where Φ is the photon flux entering the cavity and T_0 the empty-cavity transmission (*14*). The dashed lines are calculated for $r_m/\Omega = 22.5$, the value expected from the measured photon flux incident onto the cavity, $\Phi = 21 \times 10^6\ \mathrm{s}^{-1}$. The full model, including spontaneous emission (solid lines), gives very similar predictions for ρ'_{ii}.

For a given cavity, higher Φ increases the measurement rate and therefore reduces the contamination of the state by $|0_N\rangle$. However, it also increases the spontaneous emission rate and, thus, the contamination by states outside the symmetric subspace. The optimum measurement rate is a compromise between these conflicting effects. We have investigated this by varying Φ, as shown in Fig. 4. By solving the full model described above, we obtain the solid curves in Fig. 4, which are in good agreement with the experimental data and show a broad maximum of ρ_{11} as a function of r_m. The data in Fig. 2B are taken at $r_m = 22.5\Omega$, maximizing the number of entangled particles as deduced from the criterion (*15*).

By detecting transmitted photons on a photodetector behind the cavity, we can access the result of the Zeno measurement; that is, we can obtain information about whether the dynamics was restricted to Z. This additional information can be exploited to improve the fidelity of the produced state, at the price of excluding some data. In the data of Fig. 2, a transmitted photon is observed for 5 to 17% of the runs, depending on t. Excluding these runs from the analysis improves the quality of resulting state, as shown by the open symbols in Figs. 3 and 4. As an example, for trajectory I at $t = 0.96T$, we obtain an entanglement depth of 11^{+2}_{-3} atoms.

The state fidelity can be drastically improved with a better cavity, using technology that is already available today. In our cavity, mirror birefringence causes a second, orthogonally polarized transverse electromagnetic TEM00 mode located at a detuning of 540 MHz from the probed mode. This makes the spontaneous emission rate much larger than in a birefringence-free cavity with the same linewidth and g. In the latter, the spontaneous emission rate for atoms in $|n_N\rangle$ is $r_m/(2nC)$ for $n \geq 1$ (*14*). For our cavity and transition, the atomic and cavity half-linewidths at half maximum are $\gamma = 2\pi \times 3$ MHz and $\kappa = 2\pi \times 53$ MHz, so that $C \simeq 100$. Fiber cavities have been fabricated recently with a finesse approaching 200,000 (*21*–*23*) (to be compared with 37,000 in our cavity) and with a birefringence reduced to zero (*23*). With these improvements alone and all other parameters unchanged, we calculate a state fidelity $\rho_{11} \simeq 0.8$.

These results demonstrate that QZD is an experimentally feasible tool for quantum engineering of multiparticle systems. To date, many QZD proposals have focused on two-qubit systems (*24*–*29*). Further exploration involving extensions of the scheme studied here should be promising. For example, combining driven Rabi oscillations with a nondemolition measurement on the equator of the Bloch sphere (*30*, *31*) could lead to highly entangled states in the vicinity of the measurement boundary, similar to the mechanism proposed for photons in (*10*).

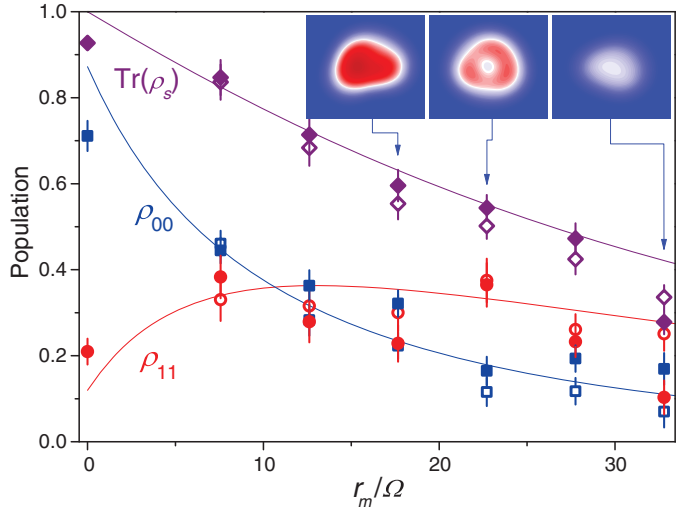


Fig. 4. QZD for different measurement strengths. Tomography measurements are taken after a fixed evolution time $t = 0.96T$ for different measurement rates (see text), using trajectory I. Solid symbols denote reconstructed ρ_{00} (blue squares) and ρ_{11} (red circles) and population in the symmetric subspace (purple diamonds). Solid lines indicate results of the full model. The insets show some of the reconstructed Husimi-Q distributions. Increasing the effective measurement rate from zero, ρ_{00} decreases whereas ρ_{11} increases as normal dynamics turns into QZD. At the same time, spontaneous emission increases, reducing the population in the symmetric subspace. Open symbols represent runs with nonzero cavity transmission excluded. Error bars are 1σ statistical errors of the reconstruction.

REFERENCES AND NOTES

1. F. Verstraete, M. M. Wolf, J. I. Cirac, *Nat. Phys.* **5**, 633–636 (2009).
2. I. D. Leroux, M. H. Schleier-Smith, V. Vuletić, *Phys. Rev. Lett.* **104**, 073602 (2010).
3. J. T. Barreiro *et al.*, *Nature* **470**, 486–491 (2011).
4. H. Krauter *et al.*, *Phys. Rev. Lett.* **107**, 080503 (2011).
5. S. Shankar *et al.*, *Nature* **504**, 419–422 (2013).
6. Y. Lin *et al.*, *Nature* **504**, 415–418 (2013).
7. R. McConnell, H. Zhang, J. Hu, S. Ćuk, V. Vuletić, *Nature* **519**, 439–442 (2015).
8. P. Facchi, S. Pascazio, *Phys. Rev. Lett.* **89**, 080401 (2002).

9. P. Facchi, S. Pascazio, *J. Phys. A Math. Theor.* **41**, 493001 (2008).
10. J. M. Raimond *et al.*, *Phys. Rev. Lett.* **105**, 213601 (2010).
11. F. Schäfer *et al.*, *Nat. Commun.* **5**, 3194 (2014).
12. A. Signoles *et al.*, *Nat. Phys.* **10**, 715–719 (2014).
13. L. Bretheau, P. Campagne-Ibarcq, E. Flurin, F. Mallet, B. Huard, *Science* **348**, 776–779 (2015).
14. J. Volz, R. Gehr, G. Dubois, J. Estève, J. Reichel, *Nature* **475**, 210–213 (2011).
15. F. Haas, J. Volz, R. Gehr, J. Reichel, J. Estève, *Science* **344**, 180–183 (2014).
16. Supplementary materials are available on *Science* Online.
17. W. Dür, G. Vidal, J. I. Cirac, *Phys. Rev. A* **62**, 062314 (2000).
18. L. Pezzé, A. Smerzi, *Phys. Rev. Lett.* **102**, 100401 (2009).
19. A. S. Sørensen, K. Mølmer, *Phys. Rev. Lett.* **86**, 4431–4434 (2001).
20. H. Strobel *et al.*, *Science* **345**, 424–427 (2014).
21. A. Muller, E. B. Flagg, J. R. Lawall, G. S. Solomon, *Opt. Lett.* **35**, 2293–2295 (2010).
22. B. Brandstätter *et al.*, *Rev. Sci. Instrum.* **84**, 123104 (2013).
23. M. Uphoff, M. Brekenfeld, G. Rempe, S. Ritter, *New J. Phys.* **17**, 013053 (2015).
24. X. B. Wang, J. Q. You, F. Nori, *Phys. Rev. A* **77**, 062339 (2008).
25. S. Maniscalco, F. Francica, R. L. Zaffino, N. Lo Gullo, F. Plastina, *Phys. Rev. Lett.* **100**, 090503 (2008).
26. R. Rossi, K. M. Fonseca Romero, M. C. Nemes, *Phys. Lett. A* **374**, 158–160 (2009).
27. X. Q. Shao, L. Chen, S. Zhang, K.-H. Yeon, *J. Phys. B* **42**, 165507 (2009).
28. C. M. Chandrashekar, *Phys. Rev. A* **82**, 052108 (2010).
29. Z. C. Shi, Y. Xia, H. Z. Wu, J. Song, *Eur. Phys. J. D* **66**, 127 (2012).
30. M. H. Schleier-Smith, I. D. Leroux, V. Vuletić, *Phys. Rev. Lett.* **104**, 073604 (2010).
31. J. Appel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10960–10965 (2009).

ACKNOWLEDGMENTS

This work was supported by the European Union Information and Communication Technologies project QIBEC (Quantum Interferometry with Bose-Einstein Condensates) (GA 284584) and by the Institut Francilien pour la Recherche sur les Atomes Froids (IFRAF). We thank J. Volz, R. Gehr, and G. Dubois for their contributions to early stages of this experiment. G.B. and L.H. performed the experiment; F.H. contributed to the experimental setup; and G.B., L.H., J.E., and J.R. contributed to data analysis and interpretation, as well as to writing the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1317/suppl/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References (*32*–*38*)

13 October 2014; accepted 18 August 2015
10.1126/science.aaa0754

CERAMIC FUEL CELLS

Readily processed protonic ceramic fuel cells with high performance at low temperatures

Chuancheng Duan,[1] Jianhua Tong,[1]* Meng Shang,[1] Stefan Nikodemski,[1] Michael Sanders,[1] Sandrine Ricote,[1] Ali Almansoori,[2] Ryan O'Hayre[1]*

Because of the generally lower activation energy associated with proton conduction in oxides compared to oxygen ion conduction, protonic ceramic fuel cells (PCFCs) should be able to operate at lower temperatures than solid oxide fuel cells (250° to 550°C versus ≥600°C) on hydrogen and hydrocarbon fuels if fabrication challenges and suitable cathodes can be developed. We fabricated the complete sandwich structure of PCFCs directly from raw precursor oxides with only one moderate-temperature processing step through the use of sintering agents such as copper oxide. We also developed a proton-, oxygen-ion–, and electron-hole–conducting PCFC-compatible cathode material, $BaCo_{0.4}Fe_{0.4}Zr_{0.1}Y_{0.1}O_{3-\delta}$ (BCFZY0.1), that greatly improved oxygen reduction reaction kinetics at intermediate to low temperatures. We demonstrated high performance from five different types of PCFC button cells without degradation after 1400 hours. Power densities as high as 455 milliwatts per square centimeter at 500°C on H_2 and 142 milliwatts per square centimeter on CH_4 were achieved, and operation was possible even at 350°C.

Among the various types of fuel cells, ceramic fuel cells possess several attractive advantages such as fuel flexibility (including the potential to directly use hydrocarbon fuels), high efficiency, and the absence of a requirement for precious-metal catalysts (*1*–*3*), but high operating temperatures [700° to 1000°C for conventional "first generation" yttria-stabilized zirconia (YSZ)–based solid oxide fuel cell (SOFCs)] result in high costs and materials compatibility challenges (*4*, *5*). "Second-generation" SOFCs (*5*, *6*), based on newer oxygen-ion–conducting electrolytes [such as samarium-doped ceria (SDC)], lowered operating temperatures to ~600°C (Fig. 1A). Nanostructured "third generation" SOFCs incorporating rare-earth elements such as Eu or Ru, and ultrathin multilayer electrolytes or core-shell nanofiber composite electrodes, have achieved exceptional performance at 450° to 600°C (*7*, *8*), but performance drops rapidly with decreasing temperature because of the high activation energy (E_a) associated with oxygen-ion conduction (Fig. 1B).

Proton conduction in oxides generally has a lower E_a compared to oxygen-ion conduction, so protonic ceramic fuel cells (PCFCs) offer intriguing potential for high-performance (*9*–*13*), lower-temperature ceramic fuel cell operation. PCFCs also offer a number of other potential benefits compared to SOFCs, particularly when operating on hydrocarbon fuels. These advantages, illustrated in Fig. 1C, include higher CH_4 conversion because of direct proton (hydrogen) removal from the anode and higher carbon coking resistance because of unfavorable Bouduard reaction. However, current PCFC performance lags far behind SOFC performance (Fig. 1A), although in the late 1990s, Kreuer *et al.* provided an important step toward enabling viable PCFCs with the demonstration of stable yttrium-doped barium zirconate (BZY) proton conductors with high (bulk) proton conductivity (*14*, *15*). Despite this advance, the high grain boundary resistance and fabrication challenges associated with this refractory material system have, until now, constrained its application. Nevertheless, the intrinsic conductivities of currently available protonic ceramic electrolytes suggest that PCFCs can eventually deliver excellent performance between 250° and 550°C. The predicted PCFC performance values based on the limits of current PCFC electrolytes (Fig. 1B) are distributed between 0.2 and 1.6 W cm^{-2} at 350° to 600°C, based on a 10-μm-thick electrolyte (similar to current third-generation SOFCs) and assuming electrode resistances identical to those reported in this study. Moreover, if epitaxial or "bamboo-structured" PCFC electrolytes can be achieved, thereby mitigating the deleterious effect of blocking grain boundaries, PCFC power densities >2.0 W cm^{-2} could be reached.

Two major reasons why PCFCs have lagged their more mature SOFC counterparts are a lack of suitable cathodes expressly designed for PCFC operation and fabrication challenges stemming from the refractory nature of most PCFC electrolytes. Here, we introduce advances that address both of these issues, leading to good PCFC performance at temperatures between 350° and 500°C with power densities of 100 to 455 mW cm^{-2}.

The poor performance of most PCFCs is attributed, in part, to their use of cathodes that were developed for SOFCs operating at much higher temperatures (700° to 1000°C) when target PCFC operation temperatures are near 500°C. We have developed a perovskite cathode composition, $BaCo_{0.4}Fe_{0.4}Zr_{0.1}Y_{0.1}O_{3-\delta}$ (BCFZY0.1), that is specifically designed for PCFCs (figs. S2 to S3) (see the supplementary materials). BCFZY0.1 is a Y-doped modification of $BaCo_{0.4}Fe_{0.4}Zr_{0.2}O_{3-\delta}$ (BCFZ), which we previously reported as a highly active and chemically compatible cathode material for

[1]Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA. [2]The Petroleum Institute, Post Office Box 2533, Abu Dhabi, United Arab Emirates.
*Corresponding author. E-mail: jhtongm@gmail.com (J.T.); rohayre@mines.edu (R.O'H.)

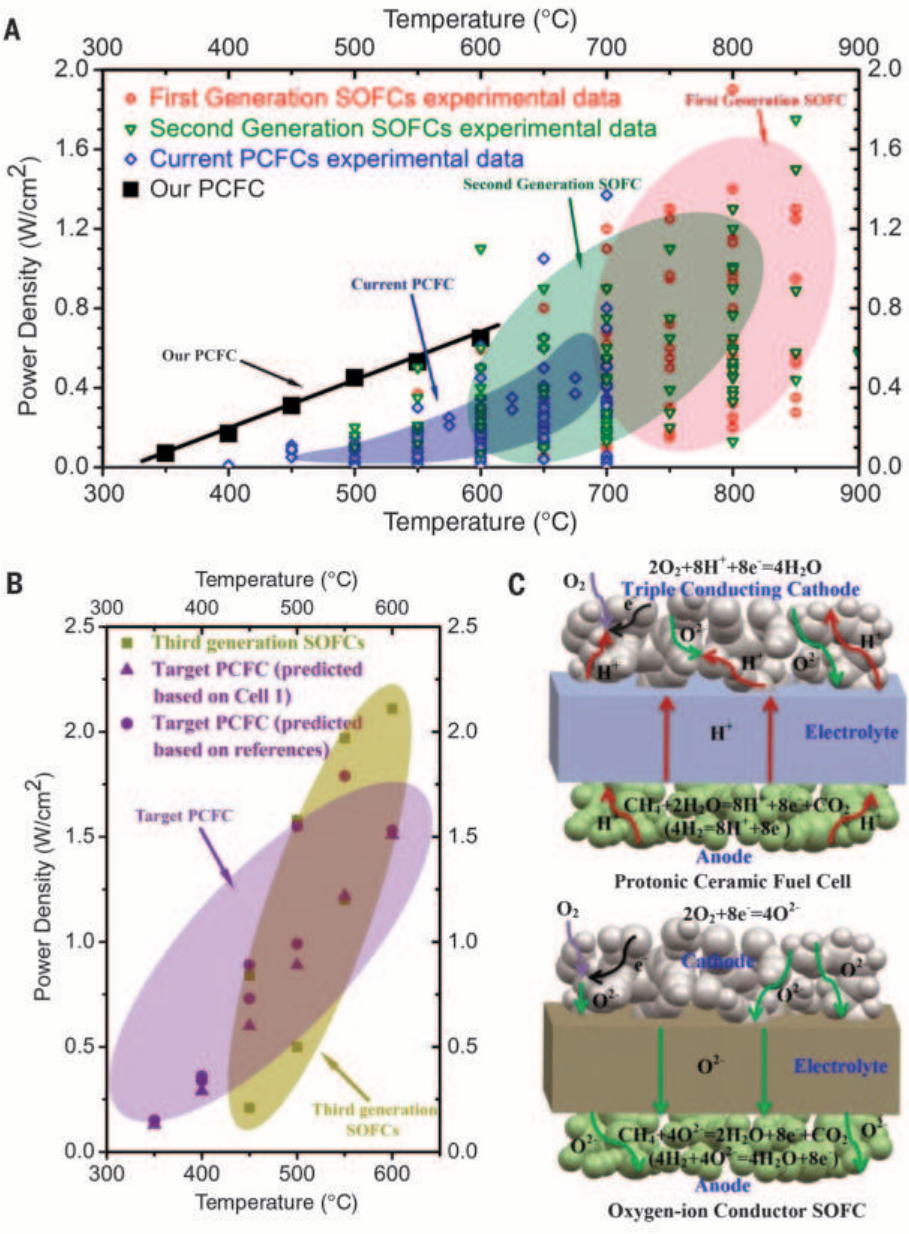


Fig. 1. Comparison of PCFCs and SOFCs. (**A**) Performance of current "first-generation" (YSZ-based) SOFCs, "second-generation" [SDC, GDC (gadolinium-doped ceria), and LSGM (strontium- and magnesium-doped lanthanum gallate)-based] SOFCs, and "first generation" PCFCs versus the SSRS-fabricated PCFC with triple-conducting oxide cathode reported here. The new SSRS-based PCFC shows excellent promise in the intermediate- and low-temperature regime (350° to 600°C). (**B**) Performance of recently reported nanostructured "third generation" SOFCs, which generally incorporate exotic elements such as Eu or Ru, ultrathin multilayer electrolytes, and/or highly nanostructured electrodes versus predicted performance of PCFCs based on the currently achievable area-specific resistance of a 10-μm-thick protonic ceramic electrolyte and assuming electrode resistances identical to those reported in this work. The predicted performance suggests that PCFCs can eventually deliver excellent performance in the IT range (250° to 550°C)—rivaling, if not surpassing, the best third-generation SOFCs. Moreover, the decreased activation energy of PCFC electrolytes compared to SOFC electrolytes suggests that PCFCs can be particularly attractive at lower temperatures. (References used for assembling the data points in Fig. 1, A and B, are provided in the supplementary materials.) (**C**) Schematic illustration of SOFC and PCFC operation. PCFCs can offer a number of other potential benefits compared to SOFCs, particularly when operating on hydrocarbon fuels. These advantages include higher CH_4 conversion because of direct proton (hydrogen) removal from the anode and higher carbon coking resistance due to conditions disfavoring the Bouduard reaction (supplementary text and fig. S1).

PCFCs (*16*, *17*). BCFZY0.1 is a transition-metal–doped derivative of the well-known proton-conducting oxide $BaZr_xY_{1-x}O_{3-\delta}$ (BZY) (*18*). Although BZY is an excellent proton conductor and also exhibits some oxygen-ion conductivity in dry reducing atmospheres (*19*), its electronic conductivity is extremely small. By heavily doping the B site of BZY with transition-metal cations (Co and Fe), the electronic percolation threshold is exceeded, thus activating electronic conduction while maintaining ionic conductivity (figs. S4 to S6). The result is a "triple conducting" cathode material (*20*, *21*) that exhibits simultaneous proton, oxygen-ion, and electron-hole conductivity (figs. S7 and S8). As illustrated in fig. S9, the application of conventional SOFC cathodes (which are based on either electron-conducting oxides or mixed oxygen-ion and electron-conducting oxides) to PCFC electrolytes restricts the cathode reaction only to points where the electrolyte and electrode phases meet. In contrast, the triple-conducting BCFZY0.1 cathode eliminates the triple-phase boundary constraints associated with traditional composite cathode architectures: The entire cathode becomes electrochemically active, which offers the chance to lower the viable operating window of PCFC devices to <400°C compared to >700°C today (table S1).

Fabrication complexity has also restrained the commercial development of PCFC technology. The basic structure of a PCFC consists of a fully dense proton-conducting ceramic electrolyte membrane sandwiched between a porous anode and a porous cathode. Traditionally (Fig. 2A), the high-quality componential powders (electrolyte, anode, and cathode) must be synthesized from expensive precursors (e.g., nitrates) by complicated wet-chemistry routes (or by time- and energy-consuming solid-state reaction procedures) followed by multiple drying, grinding, and high-temperature calcination (≥1000°C) steps. The anode support is then prepared and bisque fired, after which the electrolyte layer is deposited and the anode/electrolyte "half-cell" is cofired at temperatures higher than 1600°C to achieve acceptable electrolyte density. The high sintering temperature required to achieve densification of the protonic ceramic electrolyte generally also leads to undesirable coarsening of the anode structure. Finally, a porous cathode layer is deposited and the cell is fired a third time to complete the structure. This separate cathode deposition and firing step frequently leads to interfacial weakness between the cathode and the electrolyte and can constrain the choice of materials options.

Our PCFC fabrication method enables the full cell (i.e., porous anode, dense electrolyte, and porous cathode) to be created in a single reduced-temperature (1400°C) firing step directly from the raw precursor oxides (Fig. 2B). This approach leverages the recent development of solid-state reactive sintering (SSRS) (*10*), whereby carefully selected sintering aids can be used to assist the conversion of appropriately mixed raw precursor oxides and carbonates (e.g., $BaCO_3$ + CeO_2 + ZrO_2 +… etc.) directly into the final phase-

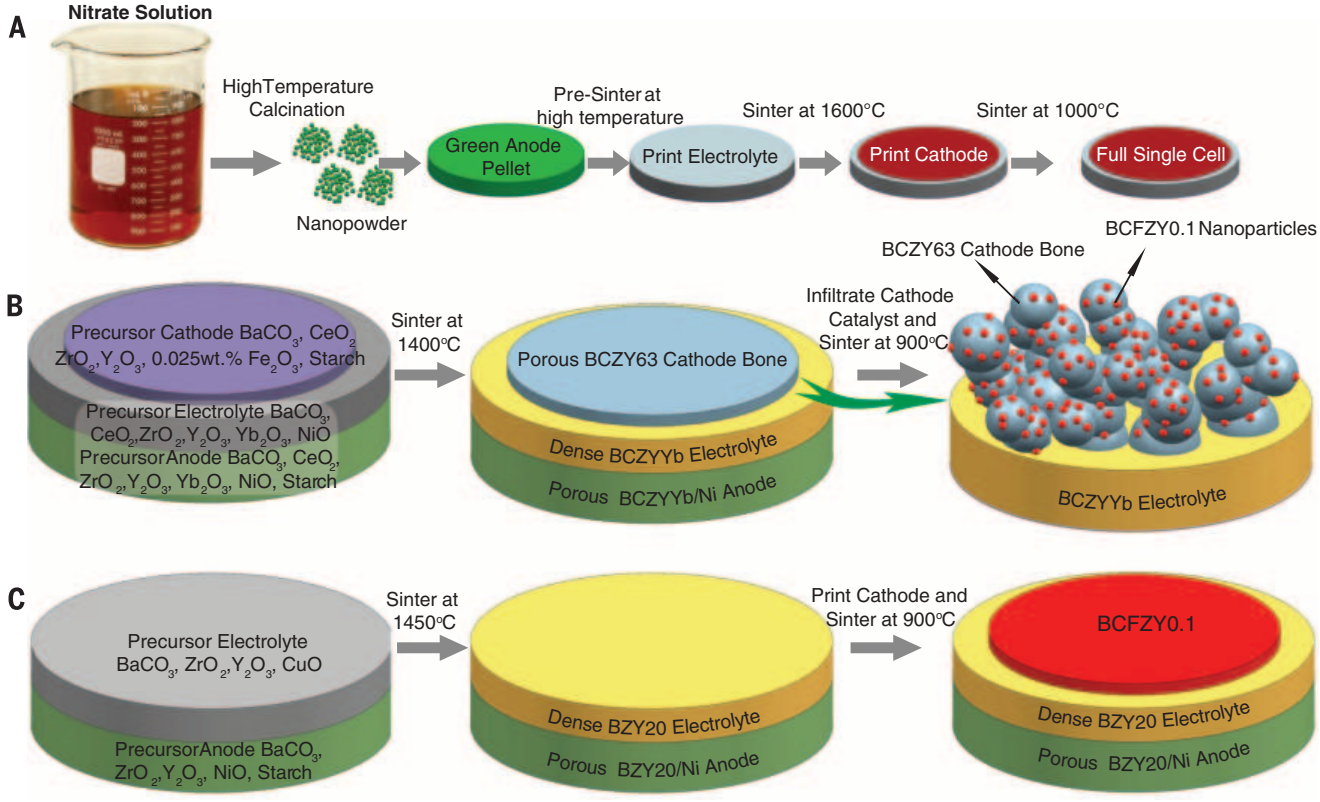


Fig. 2. Schematic illustration of the fabrication and structure of PCFC button cells. (**A**) traditional approach, (**B**) composite cathode SSRS approach, and (**C**) thin-film cathode SSRS approach.

Table 1. Fabrication method, cell composition, and peak power density of cells 1 to 5.

PCFC cells	Fabrication method	Single-cell composition	Peak power density(W cm^{-2})
Cell 1	Fig. 2B	40 wt % BCZYYb + 60 wt % NiO | BCZYYb +1.0 wt % NiO | BCZY63 + BCFZY0.1	0.455
Cell 2	Fig. 2C	40 wt % BCZYYb + 60 wt % NiO | BCZYYb + 1.0 wt % NiO | BCFZY0.1	0.405
Cell 3	Fig. 2C	45 wt % BZY20 + 55 wt % NiO | BZY20 + 1.0 wt % NiO | BCFZY0.1	0.335
Cell 4	Fig. 2C	45 wt % BZY20 + 55 wt % NiO | BZY20 + 1.4 wt % CuO | BCFZY0.1	0.268
Cell 5	Fig. 2C	40 wt % BCZY63 + 60 wt % NiO | BCZY63 + 1.3 wt % CuO | BCFZY0.1	0.318

pure anode (fig. S10), electrolyte, and cathode perovskite compositions during the single firing step. By using different sintering aids for the electrolyte versus the cathode, the former can be rendered fully dense, whereas the latter can maintain a highly porous and active nanostructure under the same sintering conditions (fig. S11). To further improve cathode performance, a second, optional step (Fig. 2B) subsequently deposits a nanoscale cathode catalyst phase into the porous cathode bone with solution infiltration followed by calcination at moderate temperatures (500° to 900°C). Alternatively, as shown in Fig. 2C, an anode + electrolyte half-cell can be sintered directly from raw precursor oxides in a first, moderate-temperature sintering step (~1400°C) with the subsequent incorporation of a single-phase thin-film cathode via a second lower-temperature (~900°C) sintering step to ensure high cathode surface area and activity.

To illustrate the versatility of this new approach, we used the SSRS method to fabricate five different types of PCFC button cells (Table 1). The button cells feature three different well-known PCFC electrolytes—$BaZr_{0.8}Y_{0.2}O_{3-\delta}$ (BZY20), $BaCe_{0.6}Zr_{0.3}Y_{0.1}O_{3-\delta}$ (BCZY63), and $BaCe_{0.7}Zr_{0.1}Y_{0.1}Yb_{0.1}O_{3-\delta}$ (BCZYYb)—in combination with two different sintering aids (CuO or NiO) and the triple-conducting (electron hole, oxygen ion, and proton) oxide BCFZY0.1 cathode. These varied cell compositions demonstrate the generality and reproducibility of our approach. BZY20, the prototypical PCFC electrolyte material, is notoriously difficult to sinter and densify. It has excellent stability, but high grain boundary resistance. BCZY63 provides improved sinterability and lower grain boundary resistance, but decreased stability compared to BZY20 (*22*, *23*). BCZYYb (*24*) demonstrates

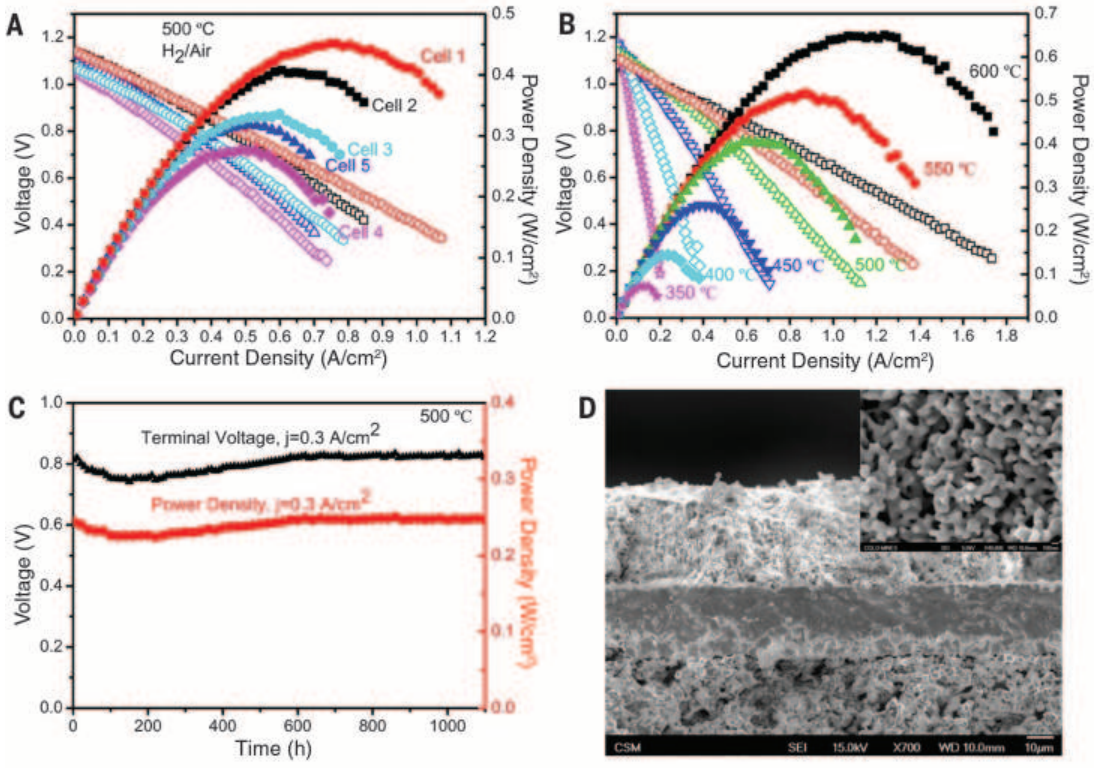


Fig. 3. Performance and microstructure of selected cells under H_2/air operation. (**A**) *I-V* and power density of cells 1 to 5 under H_2/air at 500°C; (**B**) *I-V* and power density of cell 2 under H_2/air at different temperatures; (**C**) terminal voltage and power density at a current density of 0.3 A cm^{-2} at 500°C for cell 2 under H_2/air for over 1100 hours; and (**D**) a cross-sectional view of cell 2 after operation on H_2 for over 1100 hours (inset figure is the high-magnification view of BCFZY0.1 cathode after 1100 hours operation).

one of the highest conductivities ever reported for a proton-conducting perovskite, but at the cost of further decreased stability, especially in H_2O or CO_2-containing environments (fig. S12). Nevertheless, successful fabrication of BCZYYb button cells by the single-step SSRS fabrication technique demonstrates that this approach is applicable even to compositionally complex perovskites (e.g., in this case, BCZYYb has five cations). (Experimental details on the preparation and testing of the five different PCFC button cells are provided in the supplementary materials.)

Figure 3 summarizes key results from testing of the five PCFC button cells. In Fig. 3A, the current-voltage (*I-V*) performance of all five cells is compared under H_2/air operation at 500°C. The open-circuit voltage (OCV) values for all five cells are higher than 1.05 V, suggesting that both electronic and mechanical leakages are small. Previous detailed studies of SSRS-fabricated BCZYYb–1.0 weight % (wt %) NiO electrolytes in reducing environments have demonstrated that the electronic conductivity of these electrolyte materials remains extremely small (t_e < 0.01), despite the presence of the NiO sintering aid, over a wide temperature window (100° to 800°C) (*25*). The reduced sintering temperatures enabled by our SSRS fabrication process (≤1450°C) are sufficient to fully densify the thin electrolyte layers in these cells. Figure S13 shows that dense and defect-free BZY20 electrolytes around 30 μm in thickness can be successfully fabricated with the SSRS method. All cells shown here were fabricated with 20- to 30-μm-thick electrolytes and exhibited good reliability and reproducibility.

Figure 3A reveals that cells 1 and 2, which are based on the BCZYYb electrolyte, yield the best performance, with peak power densities of 455 and 405 mW cm^{-2}, respectively, at 500°C. Previous reported power densities for PCFCs at this temperature are typically 50 to 280 mW cm^{-2} (Fig. 1A). These two BCZYYb cells differ primarily in the route used to prepare their cathodes. The highest-performing cell (cell 1) was fabricated with the route shown in Fig. 2B, whereas the other BCZYYb cell (cell 2) was fabricated with the route shown in Fig. 2C. The route 2B fabrication process produces a composite two-phase cathode with a highly porous, proton-conducting BCZY63 cathode "backbone" decorated by a nanoparticulate BCFZY0.1 catalyst phase created via a secondary infiltration process [see representative scanning electron micrograph (SEM) images in fig. S14]. In contrast, the route 2C fabrication process uses a single-phase thin-film cathode composed entirely of the BCFZY0.1 catalyst phase without a secondary proton-conducting backbone phase (see representative SEM image for this cell in Fig. 3D; additional images are provided in fig. S15). The thin-film single-phase cathode performs nearly as well as the composite cathode, which substantiates the mixed proton and electronic conduction properties of the BCFZY0.1 cathode material. BCFZY0.1 alleviates the constraints associated with traditional triple-phase boundary composite cathode architectures and enables cells to be produced by the arguably simpler route 2C fabrication process without substantial loss in performance.

Because of the fabrication advantages afforded by the simpler single-phase thin-film cathode design, cells 3, 4, and 5 were also prepared by the route 2C fabrication process. Cells 3 and 4 incorporated a BZY20 electrolyte, whereas cell 5 incorporated a BCZY63 electrolyte. Cell 3 used 1.0 wt % NiO as a sintering aid, which was mixed with the electrolyte precursors to assist in the phase-formation and densification process, whereas cells 4 and 5 used 1.4 and 1.3 wt % CuO, respectively, as a sintering aid for the same purpose. We have previously shown (*26*) that both NiO and CuO are excellent sintering aids for BZY20 and BCZY63. The BZY20 and BCZY63 cells showed modestly decreased performance compared to the BCZYYb cells, which was expected given the lower conductivity of these electrolytes. Although the electrolyte thickness and overall microstructures of cells 3 to 5 are similar (see figs. S16 to S18), the cell prepared with NiO as a sintering aid (cell 3) showed somewhat better performance. The *I-V* curves in Fig. 3A show that cell 3, with 1.0 wt % NiO as a sintering aid, has a higher OCV, which we speculate arose from a lower electronic leak compared with cell 4, which used 1.4 wt % CuO as the sintering aid.

Figure 3B provides further details on the performance of cell 2 as an example. The *I-V* performance of cell 2 as a function of temperature (Fig. 3B) shows that viable power densities (~100 mW cm^{-2}) can still be produced at temperatures as low as 350°C. Indeed, all five cells produced measurable power at 350°C (the *I-V* curves of the other four cells as a function of temperature under the same conditions are shown in figs. S16 to S19). Exemplary impedance spectroscopy plots of cells 1 and 3 are shown in fig. S20, while the electrolyte and electrode area-

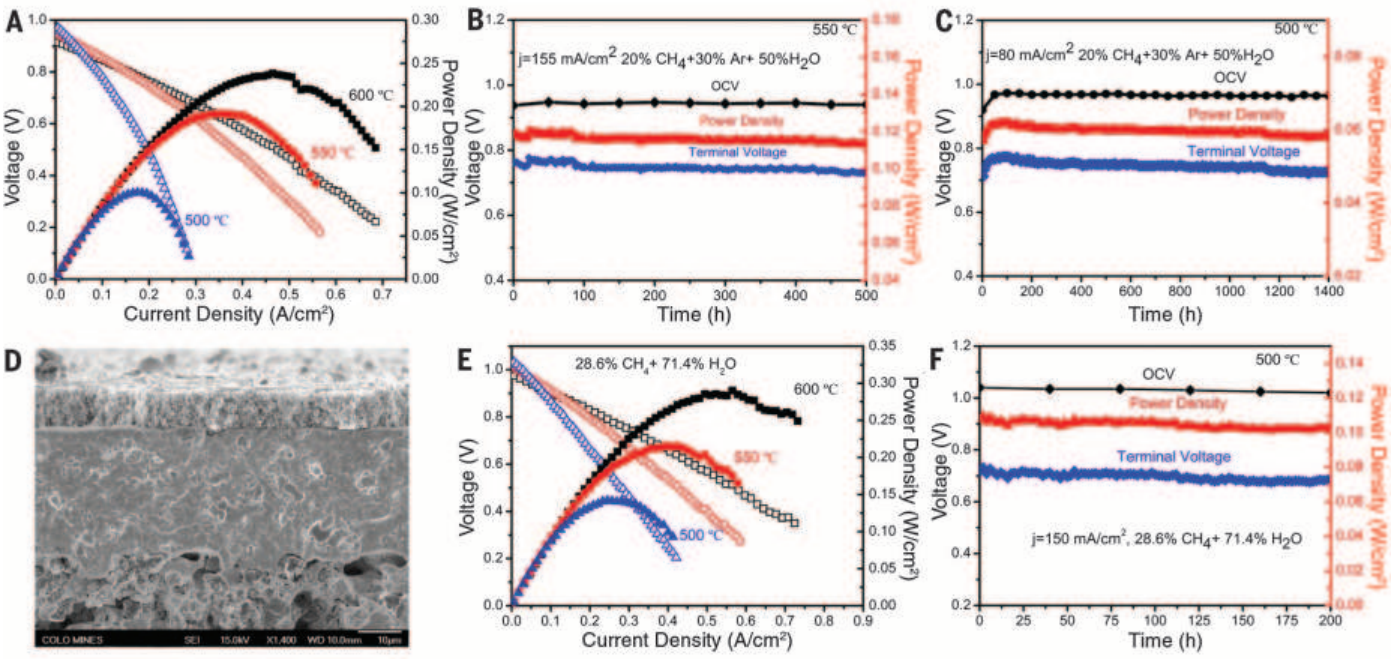


Fig. 4. Performance and microstructure of selected cells under CH_4/air operation. (**A**) *I-V* and power density for cell 4 under 20 vol. % CH_4 + 50 vol. % H_2O + 30 vol. % Ar/air at 500°, 550°, and 600°C; (**B**) terminal voltage, OCV, and power density at a current density of 155 mA cm^{-2} at 550°C for cell 4 under 20 vol. % CH_4 + 50 vol. % H_2O + 30 vol. % Ar/air for over 500 hours; (**C**) terminal voltage, OCV, and power density at a current density of 80 mW cm^{-2} at 500°C for cell 4 under 20 vol. % CH_4 + 50 vol. % H_2O + 30 vol. % Ar/air for over 1400 hours; (**D**) cross-sectional view of cell 4 after operation under 20 vol. % CH_4 + 50 vol. % H_2O + 30 vol. % Ar/air for over 1400 hours; (**E**) *I-V* and power density of cell 3 under 28.6 vol. % CH_4 + 71.4 vol. % H_2O/air at 500°, 550°, and 600°C; (**F**) terminal voltage, OCV and power density of cell 3 at a current density of 150 mA cm^{-2} at 500°C under 28.6 vol. % CH_4 + 71.4 vol. % H_2O/air for over 200 hours.

specific resistances extracted from these impedance measurements are provided in fig. S21.

Figure 3C demonstrates the stability of the operating voltage and power density during long-term testing of cell 2 under H_2/air operation at a constant current density of 300 mA cm^{-2} at 500°C. Both cell voltage and power density actually increased slightly during the course of the 1100-hour test, which we attribute to the continued reduction of the anode during the first 600 hours of operation. The cell was still fully viable after 1100 hours, and its microstructure (Fig. 3D) was virtually identical to that of an untested cell. The cathode/electrolyte and anode/electrolyte interfaces showed no signs of delamination, and the well-connected interfacial character was preserved without any visible cracking or pore formation, suggesting good thermal expansion compatibility and stability of the electrodes with the electrolyte. Furthermore, the high-magnification image of the cathode in the inset of Fig. 3D shows that even after long-term testing, the cathode maintained its fine nanostructure.

We investigated whether direct methane operation of several SSRS-fabricated fuel cells could be maintained in the intermediate-temperature operating regime. As shown in Fig. 4A, a CuO-sintered BZY20-based cell operating on direct methane fuel attains a peak power density of 240 mW cm^{-2} at 600°C [versus, e.g., 24 mW cm^{-2} at 750°C (*27*) for previous direct-methane PCFCs]. Furthermore, the cell achieved stable operation even at 500°C. The cells also maintained excellent stability. The OCV, terminal voltage, and power density for methane-fueled BZY20 fuel cells operating at 550° and 500°C remained highly stable during 500 and 1400 hours testing periods, respectively (Fig. 4, B and C). In both cases, the cells were still fully viable when testing was halted. The microstructure of the BZY20 cell after 1400 hours operation on methane at 500°C (Fig. 4D) revealed no detectable changes in cell morphology, cracking, or delamination, and no evidence of carbon deposition (fig. S22). Long-term stability under OCV conditions at 600°C (>400 hours) on methane operation was also measured for a BZY20-based fuel cell sintered with CuO (fig. S23). Based on the higher performance of the NiO-sintered BZY20 cell, its performance on methane was also tested with a H_2O/CH_4 ratio of 2.5 without fuel dilution by an inert carrier gas. The cell attained peak power densities of 290, 215, and 142 mW cm^{-2} at 600°, 550°, and 500°C, respectively (Fig. 4E). Figure 4F confirms the stability of the cell over 200 hours of testing. These direct methane PCFC single cells achieve unprecedented performance compared with previous results reported in the literature (table S2). Although BCZYYb-based cells showed better performance on hydrogen, the instability under methane operation was observed (figs. S12 and S24).

By using a densification-aiding sintering additive in the electrolyte layer, a porosity-stabilizing additive in the cathode bone, and a pore-former in the anode, solid-state reactive sintering can be used to produce a complete PCFC single cell directly from raw binary oxides with just one or two combined phase-formation and sintering steps. Low-temperature PCFC performance is further enabled by a new, triple-conducting BCFZY0.1 cathode material. The SSRS-fabricated PCFCs attain high power densities at intermediate temperature (as high as 455 mW cm^{-2} at 500°C) with viable power density produced at temperatures as low as 350°C and long-term durability of >1000 hours without loss in performance. Furthermore, SSRS-fabricated PCFCs using BZY20 electrolyte demonstrate very good intermediate-temperature performance and stability under CH_4/air testing for over 1400 hours, underscoring the promise of intermediate-temperature PCFCs for direct hydrocarbon operation. These results highlight the potential of the SSRS process to provide a commercially practical, simple, and low-cost approach to scalable solid-state ceramic devices.

REFERENCES AND NOTES

1. E. P. Murray, T. Tsai, S. A. Barnett, *Nature* **400**, 649–651 (1999).
2. S. Park, J. M. Vohs, R. J. Gorte, *Nature* **404**, 265–267 (2000).
3. T. Hibino *et al*., *Science* **288**, 2031–2033 (2000).
4. T. A. Adams II, J. Nease, D. Tucker, P. I. Barton, *Ind. Eng. Chem. Res.* **52**, 3089–3111 (2012).
5. P. I. Cowin, C. T. G. Petit, R. Lan, J. T. S. Irvine, S. Tao, *Adv Energy Mater* **1**, 314–332 (2011).
6. A. J. Jacobson, *Chem. Mater.* **22**, 660–674 (2009).
7. S. Choi *et al*., *Sci. Rep.* **3**, 2426 (2013).
8. J. G. Lee, J. H. Park, Y. G. Shul, *Nat. Commun.* **5**, 4045 (2014).
9. Y. Yamazaki, R. Hernandez-Sanchez, S. M. Haile, *Chem. Mater.* **21**, 2755–2762 (2009).
10. J. Tong, D. Clark, M. Hoban, R. O'Hayre, *Solid State Ion.* **181**, 496–503 (2010).

11. D. Pergolesi *et al.*, *Nat. Mater.* **9**, 846–852 (2010).
12. P. Babilo, T. Uda, S. M. Haile, *J. Mater. Res.* **22**, 1322–1330 (2007).
13. P. Babilo, S. M. Haile, *J. Am. Ceram. Soc.* **88**, 2362–2368 (2005).
14. K. D. Kreuer, *Solid State Ion.* **125**, 285–302 (1999).
15. K. D. Kreuer *et al.*, *Solid State Ion.* **145**, 295–306 (2001).
16. M. Shang, J. Tong, R. O'Hayre, *RSC Adv.* **3**, 15769–15775 (2013).
17. J. Tong *et al.*, *J. Membr. Sci.* **203**, 175–189 (2002).
18. K. D. Kreuer, *Annu. Rev. Mater. Res.* **33**, 333–359 (2003).
19. E. Fabbri, D. Pergolesi, E. Traversa, *Chem. Soc. Rev.* **39**, 4355–4369 (2010).
20. D. Poetzsch, R. Merkle, J. Maier, *Phys. Chem. Chem. Phys.* **16**, 16446–16453 (2014).
21. D. Hashimoto, D. Han, T. Uda, *Solid State Ion.* **262**, 687–690 (2014).
22. E. Fabbri, A. Depifanio, E. Dibartolomeo, S. Licoccia, E. Traversa, *Solid State Ion.* **179**, 558–564 (2008).
23. S. Ricote, N. Bonanos, G. Caboche, *Solid State Ion.* **180**, 990–997 (2009).
24. L. Yang *et al.*, *Science* **326**, 126–129 (2009).
25. D. Clark *et al.*, *Phys. Chem. Chem. Phys.* **16**, 5076–5080 (2014).
26. S. Nikodemski, J. Tong, R. O'Hayre, *Solid State Ion.* **253**, 201–210 (2013).
27. I. Luisetto, E. Di Bartolomeo, A. D'Epifanio, S. Licoccia, *J. Electrochem. Soc.* **158**, B1368–B1372 (2011).

ACKNOWLEDGMENTS

This work was supported by Advanced Research Projects Agency–Energy (ARPA-E) for funding under the REBELS program (award DE-AR0000493), the National Science Foundation Materials Research Science and Engineering Centers program under grant DMR-0820518, and the Petroleum Institute in Abu Dhabi, United Arab Emirates. This work is related to U.S. Patent application 62/101,285 (2015) and U.S. Patent application 14/621,091 (2015) filed by J. Tong *et al.* J.T. and R.O'H. developed the intellectual concept, designed all the experiments, and supervised this research. C.D. performed the fabrication and testing experiments of PCFC single cells. M.Sh. synthesized and tested the cathode materials. S.N. identified the copper oxide as an effective sintering aid for solid-state reactive sintering. M.Sa. measured proton concentration in cathode material. S.R. contributed to the preparation of the pastes for the electrolytes and cathode bones. A.A. participated in discussion and analysis of the methane-fueled cell testing. J.T., R.O'H., and C.D. analyzed all experimental data and wrote the paper.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1321/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Tables S1 and S2
References for Fig. 1, A and B
References (*28*–*91*)

21 April 2015; accepted 13 July 2015
Published online 23 July 2015
10.1126/science.aab3987

ORGANIC CHEMISTRY

Site-selective arene C-H amination via photoredox catalysis

Nathan A. Romero, Kaila A. Margrey, Nicholas E. Tay, David A. Nicewicz*

Over the past several decades, organometallic cross-coupling chemistry has developed into one of the most reliable approaches to assemble complex aromatic compounds from preoxidized starting materials. More recently, transition metal–catalyzed carbon-hydrogen activation has circumvented the need for preoxidized starting materials, but this approach is limited by a lack of practical amination protocols. Here, we present a blueprint for aromatic carbon-hydrogen functionalization via photoredox catalysis and describe the utility of this strategy for arene amination. An organic photoredox-based catalyst system, consisting of an acridinium photooxidant and a nitroxyl radical, promotes site-selective amination of a variety of simple and complex aromatics with heteroaromatic azoles of interest in pharmaceutical research. We also describe the atom-economical use of ammonia to form anilines, without the need for prefunctionalization of the aromatic component.

The development of catalytic procedures for the selective modification of carbon-hydrogen (C-H) bonds carries the promise of streamlined and sustainable syntheses of high-value chemicals. Direct transformation of aryl C-H bonds into carbon-carbon (C-C), carbon-oxygen (C-O), and carbon-nitrogen (C-N) bonds can provide efficient access to arenes with diverse structural properties (*1*, *2*). In particular, interest in aryl C-H amination (construction of a C-N bond from a C-H bond) is driven by the ubiquity of aryl C-N bonds in pharmaceuticals, natural products, agrochemicals, pigments, and optoelectronic materials. In contrast to the Buchwald-Hartwig (*3*, *4*) and Chan-Lam (*5*, *6*) aminations, which stand as the current preferred methods for catalytic aryl C-N bond construction, a C-H amination strategy could circumvent the need for prior functionalization of the arene as halide, triflate, or boronic acid. This synthetic advantage is augmented by the application of C-H amination to late-stage functionalization of synthetic targets, wherein libraries of complex aryl amines could be generated in a single step for medicinal chemistry screening.

Many of the recent advances in aryl C-H amination have been propelled by the ability of transition metals to activate C-H bonds. Although a regioselective addition to an arene that lacks a strong electronic or steric bias is an intrinsic challenge of aryl C-H functionalization, a number of researchers, including Buchwald and co-workers (*7*), Daugulis and co-workers (*8*), Shen and co-workers (*9*), and Nakamura and co-workers (*10*), have achieved orthoselective addition by relying on Lewis-basic substituents to direct the site of metalation. Beyond transition metal–catalyzed approaches, imidation of arenes and heteroarenes has been achieved by Sanford and co-workers in a photoredox mediated system (*11*), as well as by Chang (*12*) and DeBoef (*13*) and their respective co-workers, who employed $PhI(OAc)_2$ as an oxidant (Ph, phenyl; OAc, acetate). In these cases, regioselectivity was modest at best. Of the intermolecular C-H amination examples reported in the literature, few operate with the arene as a limiting reagent. Exceptional in this regard are the systems reported in studies led by Ritter (*14*), Baran (*15*), and Itami (*16*), yet each method appears to be exclusive to a single nitrogen coupling partner.

Taken together, this body of precedent research illustrates a number of remaining challenges in aryl C-H amination chemistry: (i) achievement of site-selective addition; (ii) extension of the nitrogen coupling partner beyond amides and imides, including the direct synthesis of primary anilines; and (iii) achievement of atom-economical and mild synthetic conditions. In this report, we describe our efforts to develop a C-H amination methodology that addresses these limitations and demonstrates the combination of organic photoredox catalysis with nitroxyl radicals as co-catalysts.

We hypothesized that an arene cation radical could serve as a key reactive intermediate in a direct, intermolecular C-H aryl amination. We believed that an amine could form σ-adduct **2** with an arene cation radical **1**, generated upon photoinduced electron transfer (PET) from the arene to an excited-state photoredox catalyst (cat*) (Fig. 1) (*17*–*21*). The subsequent deprotonation of distonic cation radical **2,** followed by oxidative aromatization of intermediate **3**, would deliver the desired aminated arene. As this process constitutes a two-electron and two-proton loss, an equivalent of a two-electron oxidant would be required for each photocatalyst turnover. In addition to an earlier report of an intramolecular cyclization initiated by PET (*22*), several recent investigations suggested that such a process was feasible. First, Yoshida and co-workers reported the synthesis of aryl amines by means of electrochemical oxidation (*23*–*25*). Essential to this achievement was the use of protected amines to insulate the C-N–coupled products from subsequent oxidative degradation. Accordingly, an additional synthetic step was required to liberate the desired targets. Second, Fukuzumi and co-workers studied the addition of bromide and fluoride anions to arene cation radicals, generated upon PET, via an organic photoredox catalyst (*26*, *27*). Dioxygen (O_2) served as a terminal oxidant and was believed to play a role in both the regeneration of the photoredox catalyst and the aromatization to furnish the aryl halide.

These studies lend support for the arene amination blueprint outlined in Fig. 1, and, given that aerobic conditions have been used in previous oxidative photoredox processes, O_2 was an attractive

Department of Chemistry, University of North Carolina–Chapel Hill, Chapel Hill, NC 27599-3290, USA.
***Corresponding author. E-mail: nicewicz@unc.edu**

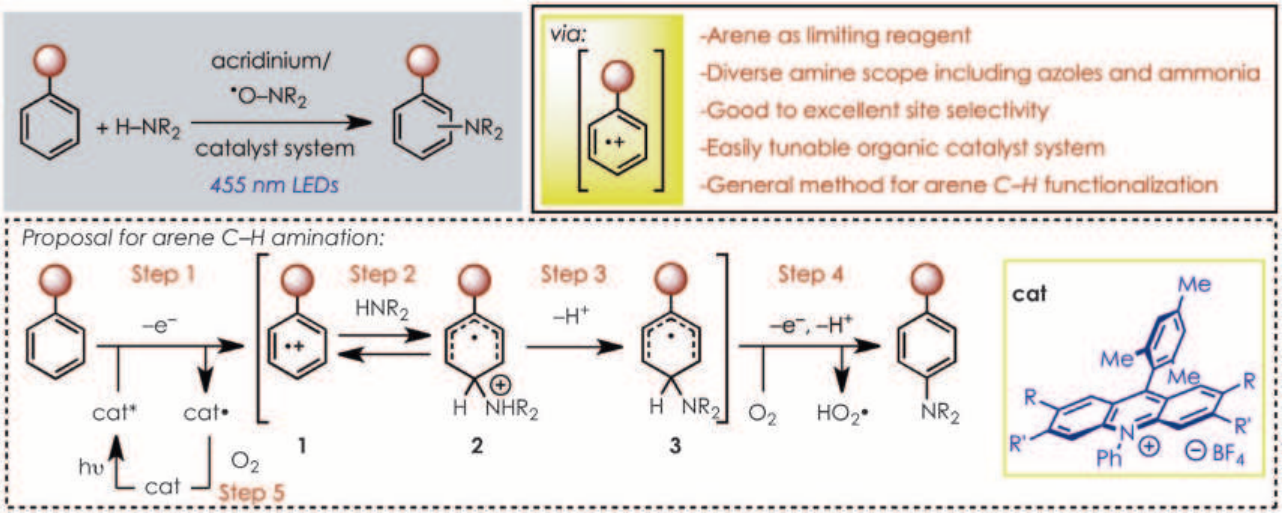


Fig. 1. A blueprint for site-selective C-H amination of aromatics. LEDs, light-emitting diodes; hυ, light.

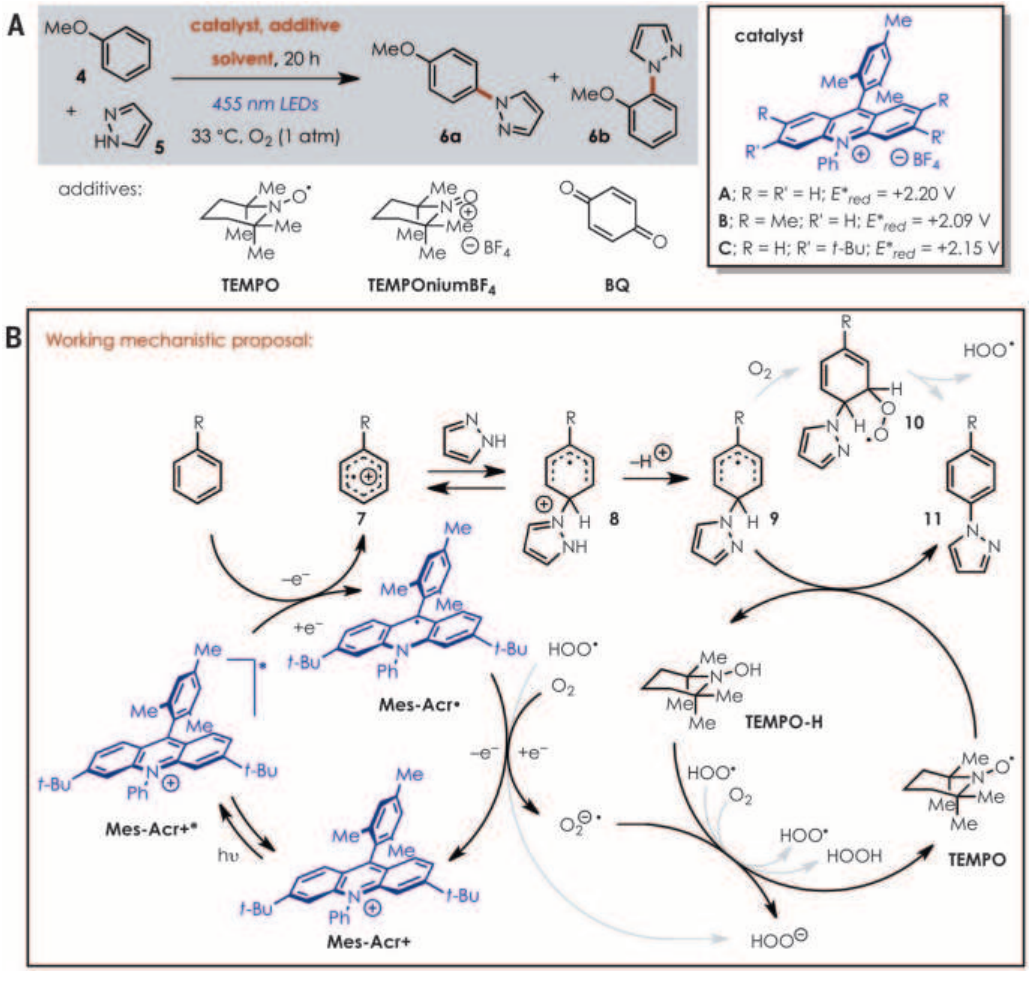


Fig. 2. Reaction development. (**A**) Catalyst optimization and (**B**) the proposed mechanism. Reactions run with 1.0 equivalent of **4** and 2.0 equivalents of **5**, unless otherwise noted. E^*_{red} values for **A** to **C** are given versus SCE (see the supplementary materials for details). BQ, 1,4-benzoquinone.

choice as a terminal oxidant and was our starting point for this investigation.

In our initial screens for reactivity, we used commercially available acridinium catalysts **A** and **B** (Fig. 2, inset), as they have highly positive excited-state reduction potentials [E^*_{red} = +2.20 and +2.09 V versus the saturated calomel electrode (SCE), respectively] and are robust in the presence of strong nucleophiles. We selected pyrazole (**5**) as a representative nucleophile and anisole (**4**) as the arene coupling partner (*28*). Under the conditions given in Fig. 2A, but in the absence of oxygen, little C-N–coupled arene adduct (**6a** and **6b**) was observed. However, when the reaction was run under a balloon of O_2, a combined 47% yield of **6a** and **6b** was observed, with good para:ortho selectivity (ratio of 6.7:1). Subsequent first-pass optimization efforts produced

no gain in yield for the catalyst, concentration, solvent, or other oxidants.

This plateau in yield could have several causes. First, aryl amine products **6a** and **6b** (irreversible half-peak potential, $E_{p/2}$ = +1.50 V versus SCE) possess lower oxidation potentials than anisole does ($E_{p/2}$ = +1.87 V versus SCE), and **6a** and **6b** could competitively reduce excited-state acridinium (cat^{+*}), resulting in product inhibition. Second, analysis of the reaction mixture revealed that phenyl formate was the major byproduct, indicating that, in addition to product inhibition, side reactions of the arene reactant were problematic under these conditions. Third, after failing to detect catalyst **A** or **B** in crude proton nuclear magnetic resonance (^{1}H NMR) spectra, we questioned the stability of the catalyst under the reaction conditions. Moreover, both anisole (**4**) and acridinium are susceptible to degradation reactions in the presence of oxygen-centered radicals (*29*); we therefore surveyed a number of additives that could mitigate any highly reactive radical intermediates, such as peroxyl radicals.

We found that 10 mol % 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) improved the yield of **6a** and **6b** to 65%. We also observed that the remaining mass balance was almost entirely unreacted anisole. Increased equivalents of TEMPO afforded a yield of 74% that decreased with higher loadings.

As an additional measure to prolong the viability of the acridinium catalyst, we modified the acridinium structure to confer stability against addition of nucleophiles or radicals (9-mesityl-3,6-di-*tert*-butyl-10-phenylacridinium tetrafluoroborate, **C**). The use of this catalyst provided the best results to date, producing compound **6** in 88% yield after 20 hours. A 97% yield was achieved under an atmosphere of air after irradiation for 3 days. The use of immobilized TEMPO on polystyrene resulted in a 65% yield of the aminated arene and facilitated its recovery and reuse via simple filtration.

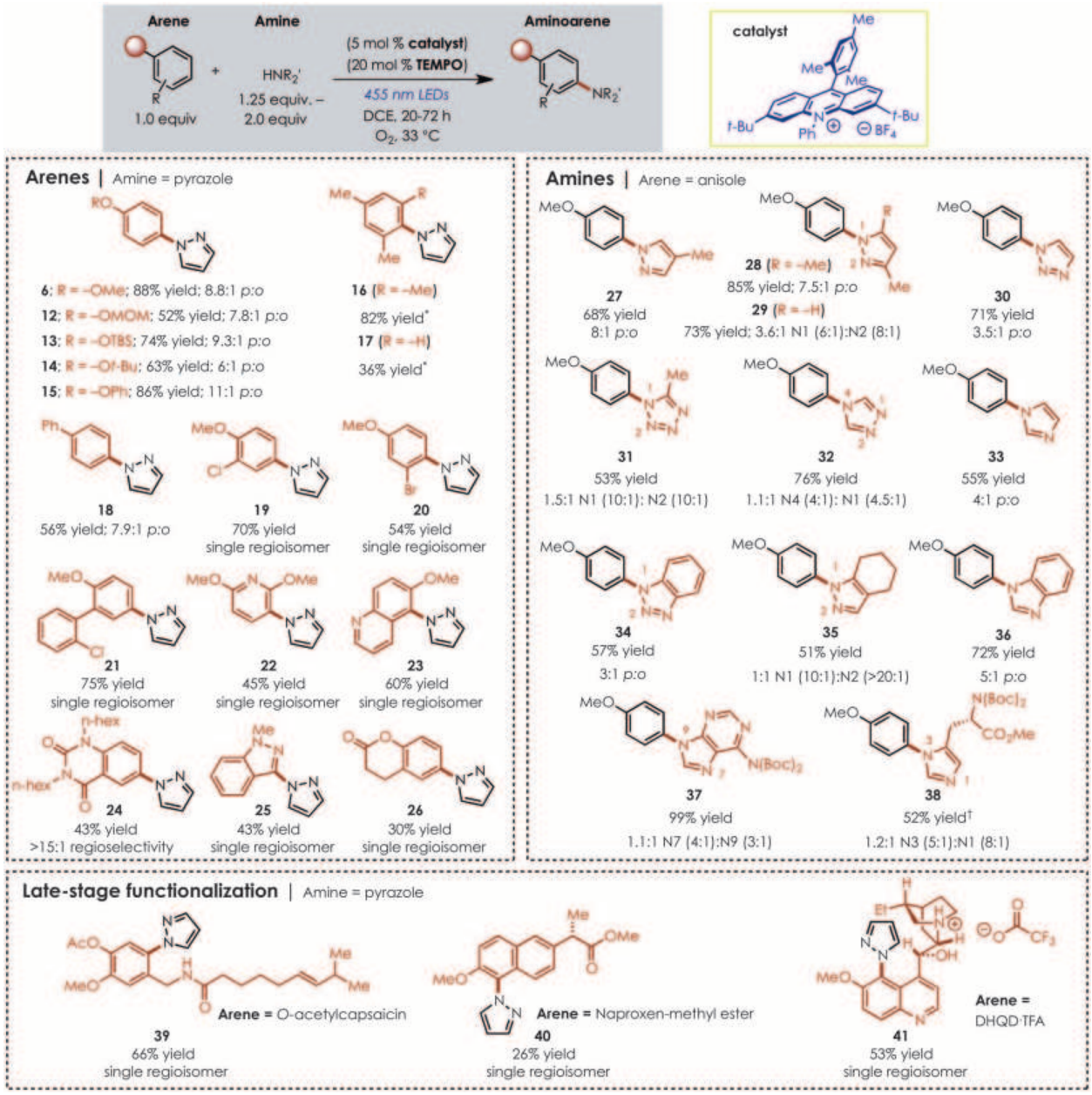


Fig. 3. Reaction scope for the C-H amination. Parenthetical ratios refer to para:ortho (p:o) selectivity for the given *N*-isomer. Reactions were run in 1,2-dichloroethane (DCE) at 0.1 M concentration with respect to the arene limiting reagent. The asterisk indicates a reaction run with 2.0 equivalents of arene, 1.0 equivalent of amine, and 1.0 equivalent of TEMPO under an N_2 atmosphere for 44 hours. The dagger indicates a reaction run under N_2 with 1.0 equivalent of TEMPO. Hex, hexyl group; Ac, acetyl group.

The mechanism of this reaction is currently the subject of detailed investigation. We believe the role of TEMPO is to aromatize radical intermediate **9** directly by H-atom abstraction (Fig. 2B). Alternatively, radical **9** could be trapped by O_2 to form 1,3-cyclohexadienyl peroxyl radical **10**, from which internal elimination would yield product **11** and hydroperoxyl radical $HO_2^{\bullet}$ (*30*). As proposed in (*26*), O_2 can oxidize acridine radical Mes-Acr• (Mes, mesityl; Acr, acridinium), regenerating acridinium Mes-Acr+ and superoxide $O_2^{-\bullet}$, although other putative intermediates might be capable of catalyst turnover (e.g., HOO•) (Fig. 2B). The strongly basic superoxide should deprotonate intermediate **10**, then undergo hydrogen atom transfer with TEMPO-H, ultimately forming H_2O_2 and regenerating TEMPO. The decrease in undesired byproducts observed when TEMPO was included is consistent with the proposed activity of TEMPO-H, which is expected to scavenge reactive oxygen-centered radicals, such as hydroperoxyl radical $HO_2^{\bullet}$. Although the half-wave redox potential of TEMPO [$E_{1/2}$ (TEMPO•/TEMPO$^+$) = +0.62 V versus Ag/AgCl] (*31*) points to the possibility of oxidization by cat^{+}*, the use of 20 mol % TEMPOnium-BF_4 produced comparable results to TEMPO in the aryl amination reaction (table S1). This suggests that a common mechanistic intermediate is accessible—namely, TEMPO—presumably generated by electron transfer from cat• ($E_{1/2}$ (cat$^+$/cat•) = −0.47 to −0.58 V versus SCE) to TEMPOnium. In the absence of cat$^+$, none of aryl amine **11** was generated with 20 mol % TEMPO, although trace product formation was detected when 20 mol % TEMPOnium-BF_4 was used and the acridinium photocatalyst was omitted.

The optimized conditions were successfully extended to the coupling of pyrazole with a variety of monosubstituted aromatics, including CH_2OCH_3 (MOM)– and *tert*-butyldimethylsilyl (TBS)–protected phenol as well as biphenyl (**12** to **15, 18**; Fig. 3). Halogenated anisole derivatives were excellent substrates for the transformation and afforded *N*-arylpyrazoles **19** and **20**, with complete regioselectivity para to the methoxy substituent. Likewise, regiochemical discrimination is possible on biaryls bearing electronically distinct aromatic groups. Despite the availability of eight unique aryl C-H bonds in 2-chloro-2'-methoxy-1,1'-biphenyl, biaryl **21** was formed in 75% yield, with completely site-selective addition para to the methoxy group, reflective of the electronic influences on this manifold. Heterocycles bearing electron-releasing substitution are competent substrates: Dimethoxypyridine **22** and methoxyquinoline **23** were isolated in modest yields but as single products. Heterocyclic motifs such as quinazoline dione, 1-methyl indazole, and dihydrocoumarin readily underwent C-H amination with pyrazole to produce adducts **24** to **26**. In all cases, a regioselectivity ratio of >15:1 was observed.

One of the challenges associated with the oxidative functionalization of arenes is the presence of weak benzylic C-H bonds, particularly in arene cation radicals, which have a documented propensity for H-atom and/or proton loss at these positions (*32*). For example, under the electrochemical oxidation conditions in (*24*), alkyl-substituted arenes give rise to benzylic amination over aryl amination. Our initial attempts to apply the previously optimized conditions to the coupling of pyrazole with mesitylene were hampered by competitive benzylic oxidation to the aryl aldehyde (table S2), a reactivity previously documented in (*33*). Excluding O_2 suppressed benzylic oxidation and increasing the TEMPO loading to 1.0 equivalent enabled the addition of pyrazole to the aromatic ring of mesitylene, forming **16** in excellent yield (82%). No products resulting from benzylic oxidation were observed. Likewise, *m*-xylene reacted under these conditions, albeit in lower yields (36%); the remainder of the mass balance was attributed to unreacted starting material. Even modest yields are notable in this context, given the oxidation potential of *m*-xylene ($E_{p/2}$ = +2.28 V versus SCE) and the excited-state reduction potential of catalyst **C**. Considering the acidity of alkylbenzene cation radicals [pK_a [PhMe]$^{+\bullet}$ = −20, where K_a is the acid dissociation constant (*34*)], it is remarkable that productive aryl C-H amination occurs for mesitylene and *m*-xylene.

Azoles are a privileged structural unit in pharmacologically active compounds (*35*, *36*) and in the architectures of transition metal–catalysts and organocatalysts. Yet the most reliable methods for constructing aryl-azoles require at least two synthetic steps. We found that a diverse range of *N*-heterocyclic nucleophiles could be directly coupled to an arene in our reaction protocol. In addition to pyrazoles (**27** to **29**), we found that 1,2,3- and 1,2,4-triazoles (**30**, **32**), tetrazole (**31**), imidazole and benzimidazole (**33** and **36**), benzotriazole (**34**), and tetrahydro-indazole (**35**) produced good to excellent yields of the C-N adducts (53 to 85%). A di-Boc–protected adenine (Boc, butoxycarbonyl) gave nearly quantitative yields (99%) of purines (**37**) in a 1.1:1 *N*-regioisomeric ratio.

To evaluate whether this catalyst system could be applied to late-stage functionalization, we tested the C-N bond–forming protocol with representative druglike molecules, as shown in Fig. 3 (bottom). The successful coupling of Boc-histidine methyl ester with **4** offers a new strategy for the modification of biologically relevant structures containing this amino acid. When reacted with pyrazole, *O*-acetylcapsaicin, naproxen methyl ester, and dihydroquinidine·trifluoroacetic acid (DHQD·TFA) were transformed into single regioisomers of the adducts (**38** to **41**). Despite heteroatom substitution at the benzylic position, no oxidation of the benzylic C-H bonds was observed in either *O*-acetylcapsaicin or DHQD·TFA in the reactions forming **39** and **41**, respectively. Likewise, naproxen methyl ester contains a sensitive benzylic C-H bond that remained undisturbed in the coupling reaction. These results demonstrate the mildness and practicality of the protocol.

The regioselectivities observed in these transformations are challenging to interpret, given the diversity of substituents on the arene coupling partner. Previous studies have found qualitative correlations between the observed site selectivity and the lowest unoccupied molecular orbital coefficients (*23*) or partial atomic charges (*26*). The aforementioned work is consistent with the expectation of nucleophilic addition to a cation radical at positions that afford a stabilized radical; in arenes bearing a single substituent, addition at the ortho and para positions is favored over meta-addition. Other differentiating factors, such as steric effects, may be intertwined with arene electronics, and future mechanistic studies could clarify the key contributions to the regioselectivities observed.

Last, we explored whether anilines could be forged directly from this catalytic sequence by using either ammonia or an ammonium salt as the nitrogen source. Traditionally, a nitration-hydrogenation sequence is used to access anilines directly. The latter protocol requires rigorous

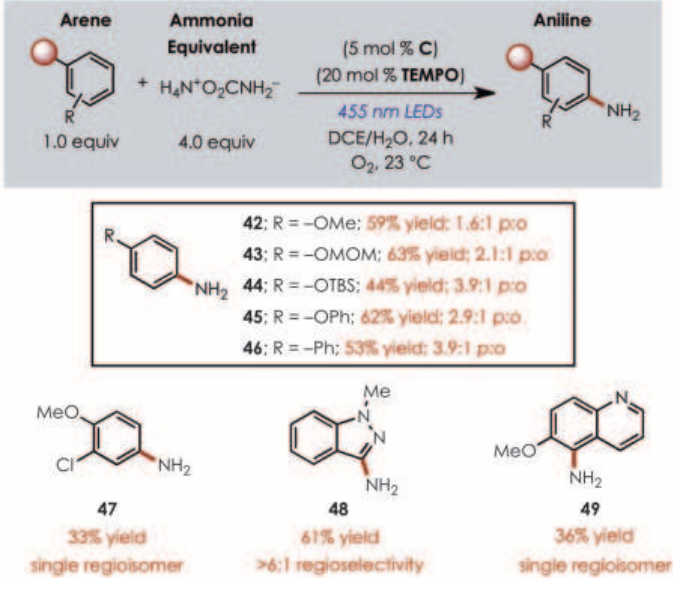


Fig. 4. Synthesis of anilines using ammonium salt as ammonia equivalent. Reactions were run in DCE and H_2O (10:1) at 0.1 M concentration with respect to the arene limiting reagent.

optimization to ensure safe dissipation of the heat associated with the exothermic reaction profile; potentially explosive intermediates and toxic byproducts are also concerns. Only recently has the Buchwald-Hartwig amination of aromatic halides been accomplished with ammonia as the nitrogen source (*37*). A C-H amination protocol of benzene with ammonia, developed by DuPont, uses a NiO-ZrO_2 catalyst system at 350°C and 300 to 400 atm, producing aniline in a 14% maximum yield (*38*, *39*).

After screening a variety of commercially available ammonium salts such as $H_4N^+OAc^-$, $H_4N^+HCO_3^-$, and $(H_4N^+)_2CO_3^{2-}$, we found that ammonium carbamate ($H_4N^+H_2NCO_2^-$) was best suited for this role (table S3 and supplementary materials). This benchtop-stable solid salt is less costly on a molar basis than liquid ammonia. Using 4.0 equivalents of ammonium carbamate with anisole, under catalytic conditions nearly identical to those applied to azoles, resulted in the formation of a 1.6:1 mixture of *para*- and *ortho*-anisidine in 59% isolated yield (**42**; Fig. 4).

The scope of the aniline-forming reaction was similar to the azole-coupling transformations. Protected phenols (**43** to **45**), haloarenes (**47**), and nitrogen heteroaromatics such as *N*-methylindazole (**48**) and 6-methoxyquinoline (**49**) were aminated under this protocol, albeit with modest regioselectivities in the case of the monosubstituted aromatics.

Overall, these C-N bond–forming reactions are powerful tools for the synthesis of complex aromatics using an organic photooxidant and nitroxyl radical catalyst system. From the substrate scope investigation, it is clear that free alcohols, esters, silyl ethers, halides, amides, alkenes, and protected amines are all compatible functionalities. The mildness of this protocol makes it appealing for a variety of applications. Moreover, we anticipate that this general method for the activation of arenes will result in the development of additional transformations.

REFERENCES AND NOTES

1. T. W. Lyons, M. S. Sanford, *Chem. Rev.* **110**, 1147–1169 (2010).
2. G. B. Shul'Pin, in *Transition Metals for Organic Synthesis: Building Blocks and Fine Chemicals*, M. Beller, C. Bolm, Eds. (Wiley-VCH, New York, ed. 2, 2004), pp. 215–241.
3. J. P. Wolfe, S. Wagaw, J.-F. Marcoux, S. L. Buchwald, *Acc. Chem. Res.* **31**, 805–818 (1998).
4. J. F. Hartwig, *Acc. Chem. Res.* **31**, 852–860 (1998).
5. P. Y. S. Lam, G. Vincent, C. G. Clark, S. Deudon, P. K. Jadhav, *Tetrahedron Lett.* **42**, 3415–3418 (2001).
6. K. Sanjeeva Rao, T.-S. Wu, *Tetrahedron* **68**, 7735–7754 (2012).
7. W. C. P. Tsang, R. H. Munday, G. Brasche, N. Zheng, S. L. Buchwald, *J. Org. Chem.* **73**, 7603–7610 (2008).
8. L. D. Tran, J. Roane, O. Daugulis, *Angew. Chem. Int. Ed.* **52**, 6043–6046 (2013).
9. H. Xu, X. Qiao, S. Yang, Z. Shen, *J. Org. Chem.* **79**, 4414–4422 (2014).
10. T. Matsubara, S. Asako, L. Ilies, E. Nakamura, *J. Am. Chem. Soc.* **136**, 646–649 (2014).
11. L. J. Allen, P. J. Cabrera, M. Lee, M. S. Sanford, *J. Am. Chem. Soc.* **136**, 5607–5610 (2014).
12. H. J. Kim, J. Kim, S. H. Cho, S. Chang, *J. Am. Chem. Soc.* **133**, 16382–16385 (2011).
13. A. A. Kantak, S. Potavathri, R. A. Barham, K. M. Romano, B. DeBoef, *J. Am. Chem. Soc.* **133**, 19960–19965 (2011).
14. G. B. Boursalian, M.-Y. Ngai, K. N. Hojczyk, T. Ritter, *J. Am. Chem. Soc.* **135**, 13278–13281 (2013).
15. K. Foo, E. Sella, I. Thomé, M. D. Eastgate, P. S. Baran, *J. Am. Chem. Soc.* **136**, 5279–5282 (2014).
16. T. Kawakami, K. Murakami, K. Itami, *J. Am. Chem. Soc.* **137**, 2460–2463 (2015).
17. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322–5363 (2013).
18. J. M. R. Narayanam, C. R. J. Stephenson, *Chem. Soc. Rev.* **40**, 102–113 (2011).
19. D. A. Nicewicz, T. M. Nguyen, *ACS Catal.* **4**, 355–360 (2014).
20. Q. Qin, S. Yu, *Org. Lett.* **16**, 3504–3507 (2014).
21. E. Brachet, T. Ghosh, I. Ghosh, B. König, *Chem. Sci.* **6**, 987–992 (2015).
22. G. Pandey, M. Sridhar, U. T. Bhalerao, *Tetrahedron Lett.* **31**, 5373–5376 (1990).
23. T. Morofuji, A. Shimizu, J. Yoshida, *J. Am. Chem. Soc.* **135**, 5000–5003 (2013).
24. T. Morofuji, A. Shimizu, J. Yoshida, *J. Am. Chem. Soc.* **136**, 4496–4499 (2014).
25. T. Morofuji, A. Shimizu, J. Yoshida, *Chem. Eur. J.* **21**, 3211–3214 (2015).
26. K. Ohkubo, K. Mizushima, R. Iwata, S. Fukuzumi, *Chem. Sci.* **2**, 715–722 (2011).
27. K. Ohkubo, A. Fujimoto, S. Fukuzumi, *J. Phys. Chem. A* **117**, 10719–10725 (2013).
28. T. M. Nguyen, N. Manohar, D. A. Nicewicz, *Angew. Chem. Int. Ed.* **53**, 6198–6201 (2014).
29. W. P. Hess, F. P. Tully, *J. Phys. Chem.* **93**, 1944–1947 (1989).
30. X.-M. Pan, M. N. Schuchmann, C. von Sonntag, *J. Chem. Soc. Perkin Trans. 2* **1993**, 289–297 (1993).
31. J. E. Baur, S. Wang, M. C. Brandt, *Anal. Chem.* **68**, 3815–3821 (1996).
32. M. Schmittel, A. Burghart, *Angew. Chem. Int. Ed. Engl.* **36**, 2550–2589 (1997).
33. K. Ohkubo *et al.*, *Chem. Commun. (Cambridge)* **46**, 601–603 (2010).
34. F. G. Bordwell, J. P. Cheng, *J. Am. Chem. Soc.* **111**, 1792–1795 (1989).
35. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
36. T. J. Ritchie, S. J. F. Macdonald, S. Peace, S. D. Pickett, C. N. Luscombe, *Med. Chem. Comm.* **3**, 1062–1069 (2012).
37. G. D. Vo, J. F. Hartwig, *J. Am. Chem. Soc.* **131**, 11049–11061 (2009).
38. T. W. Del Pesco, U.S. Patent 4031106 (1977).
39. T. W. Del Pesco, U.S. Patent 4001260 (1977).

ACKNOWLEDGMENTS

Financial support was provided by the David and Lucile Packard Foundation, Merck, and an Amgen Young Investigator Award. N.A.R is grateful for an NSF Graduate Fellowship, and K.A.M. was supported by a Francis Preston Venable Graduate Fellowship. A provisional patent has been filed on the methods presented here (U.S. patent application no. 62/170,632).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1326/suppl/DC1
Materials and Methods
Tables S1 to S4
References (*40*–*73*)
NMR Spectra

9 July 2015; accepted 19 August 2015
10.1126/science.aac9895

MINERAL SURFACES

X-ray–driven reaction front dynamics at calcite-water interfaces

Nouamane Laanait,[1,2]*† Erika B. R. Callagon,[2,3] Zhan Zhang,[4] Neil C. Sturchio,[5] Sang Soo Lee,[1] Paul Fenter[1]*

The interface between minerals and aqueous solutions hosts globally important biogeochemical processes such as the growth and dissolution of carbonate minerals. Understanding such processes requires spatially and temporally resolved observations and experimental controls that precisely manipulate the interfacial thermodynamic state. Using the intense radiation fields of a focused synchrotron x-ray beam, we drove dissolution at the calcite/water interface and simultaneously probed the dynamics of the propagating reaction fronts using surface x-ray microscopy. Evolving surface structures were controlled by the time-dependent solution composition, as characterized by a kinetic reaction model. At extreme disequilibria, we observed the onset of reaction front instabilities with velocities of > 30 nanometers per second. These instabilities serve as a signature of transport-limited dissolution of calcite under extreme disequilibrium.

Calcium carbonate precipitates abiotically and is synthesized by living organisms into complex and functional biomineral architectures (*1*). Combined, calcium carbonate minerals constitute a major fraction of Earth's upper crust in the form of carbonate rocks (*2*). Characterizing the rapidly evolving morphology of calcium carbonate during growth (*3*, *4*) and dissolution (*5*, *6*) is central to both a fundamental understanding of its reactivity and manipulation of its versatile functionality. The morphology of calcium carbonate phases can be imaged in situ with electron (*7*, *8*) and x-ray microscopies; however, the large radiation doses deposited by these probes can substantially alter the state of the system (*9*).

We used a focused x-ray beam to both observe and drive dissolution in a quantifiable manner (*10*). The synchrotron x-ray beam induces acidification and depletion of carbonate ions within the solution, which controlled the interfacial

[1]Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, IL, USA. [2]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [3]Department of Earth and Environmental Sciences, University of Illinois at Chicago, Chicago, IL, USA. [4]X-ray Science Division, Argonne National Laboratory, Argonne, IL, USA. [5]Department of Geological Sciences, University of Delaware, Newark, DE, USA.
***Corresponding author. E-mail: laanaitn@ornl.gov (N.L.); fenter@anl.gov (P.F.)** †Present address: Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA.

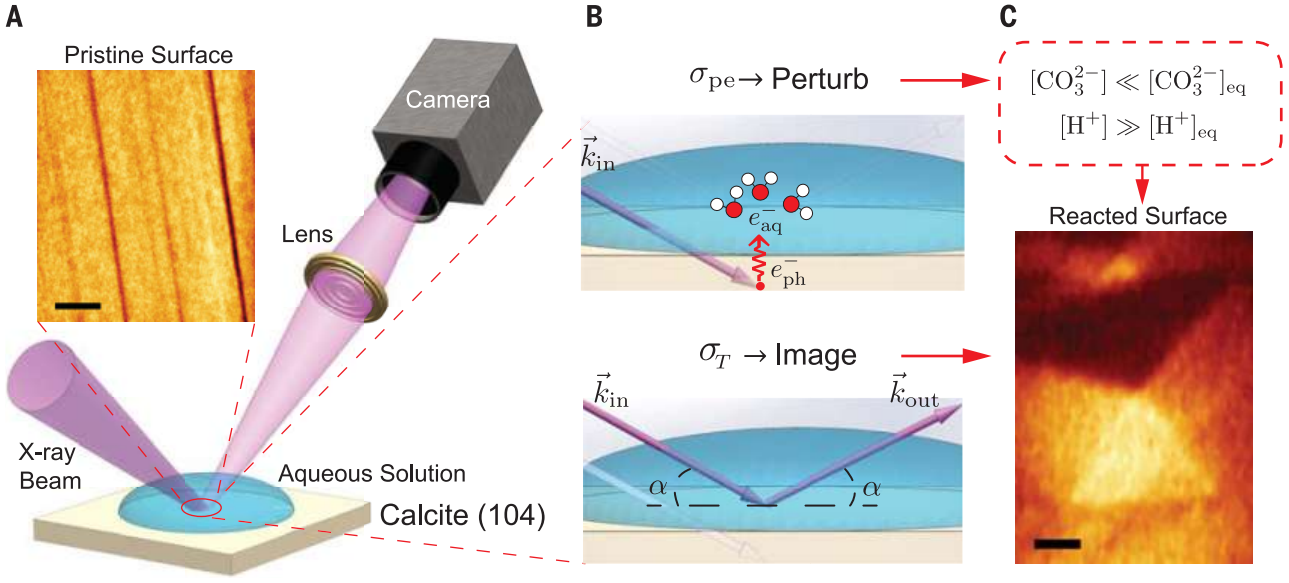


Fig. 1. Imaging the calcite/water interface while simultaneously driving it far from equilibrium. (**A**) Optical configuration of the x-ray reflection interface microscope. (Inset) XRIM image of steps on a pristine calcite surface. (**B**) (Top) Photoelectron (e^-_{ph}) generated by absorption of an incident x-ray photon in the calcite crystal. The photoelectron propagates to the mineral/water interface and disrupts the local equilibrium through the formation of short-lived radicals and hydrated electrons (e^-_{aq}). (Bottom) Elastically reflected x-rays image the calcite surface topography, with an angle of incidence α and wave vector $\mathbf{k}_{in}$. The imaging was performed at a scattering condition, $Q = 2\,|\mathbf{k}_{in}|\,\sin(\alpha) = 2.1$ Å^{-1}. (**C**) Images of etch pits formed in response to solution undersaturation. All XRIM images were flat-field–corrected and scaled to correct for distortions due to the viewing angle of the lens with respect to the crystal surface (*10*). Scale bars, 2 μm.

saturation state of solutions with respect to calcite. Simultaneously, we imaged the evolution of the calcite/water interface topography with lateral resolution below 100 nm and <1 nm height sensitivity by using x-ray reflection interface microscopy (XRIM) (*11*, *12*). We interpreted the driven morphological changes by calculating the solution saturation within a reaction kinetics model, explaining both the observed rates of mineral reaction and the dissolution mechanisms.

XRIM captured the evolving surface topography of a calcite(104) surface in contact with a thin film of calcite-equilibrated solution (about 2 μm thick) (Fig. 1). The initial surface topography (Fig. 1A and figs. S1 and S2) was a nominally flat surface with widely spaced steps. In situ imaging was enabled by the high penetration depth of 10-keV x-rays and their scattering from the mineral/water interface (the Thomson cross-section, σ_T). The absorption of x-rays by the photoelectric effect (σ_{pe}) resulted in acidification and substantial (almost 100,000-fold) undersaturation of the solution with respect to calcite, driving calcite to dissolve by multiple modes, such as etch pit development and step retreat (Fig. 1C and figs. S2 and S3). The degree of undersaturation subsequently decreased to near ~10-fold undersaturation as the system approached a steady state.

The onset of x-ray–induced dissolution was localized to the illuminated surface area. The initial unperturbed calcite surface was largely devoid of topographical features [time (t) = 0 s in Fig. 2A]. After irradiation, topographic changes began near a preexisting inhomogeneity (site 1). Subsequently, this region served as a nucleation center of a rhombic pit (site 2, t = 156 s), indicating the possible presence of a strain field in the crystal. At longer exposures, dissolution pits formed in apparently defect-free areas (homogenously nucleated pits) (sites 3 and 3′), and etch pits driven by extended defects that displayed anisotropic shapes occurred throughout the surface. Pit interactions such as annihilation due to surface retreat (site 5) and coalescence (site 4) were pronounced. Area measurements of sites 3 and 4 over time (movie S1) show the initial rapid lateral expansion of both pits and their subsequent stagnation at t >1500 s (Fig. 2B). The functional variation in pit area over time for these two sites was nearly identical except for an overall scale factor.

The occurrence of calcite etch pits is interpreted within the pit nucleation model as indicative of mineral dissolution far from equilibrium (*2*). The pit nucleation free energy, ΔG (*6*), is

$$\Delta G \sim \frac{a^3}{V_{uc}} k_B T \ln(c_{CaCO_3}/c_{eq}) + a^2\gamma - U_{elastic} \quad (1)$$

where a is the size of the pit, V_{uc} is the volume of a calcite unit cell, k_BT is the thermal energy, c_{CaCO_3} is the solution concentration of calcium carbonate, c_{eq} is the equilibrium concentration, γ is the surface energy of the crystal/water interface, and $U_{elastic}$ is the elastic energy associated with a preexisting strain field in the crystal lattice. Pit formation results as a competition between the energy cost to create new surface area (the second term in Eq. 1) and the energy gain due to crystal dissolution (first term) and lattice strain (third term). When the system is far from equilibrium ($c_{CaCO_3} \ll c_{eq}$), pit formation is expected to occur by homogeneous nucleation at point defects, as exemplified by sites 3 and 3′ (Fig. 2A). Strain distribution associated with an extended dislocation, $U_{elastic}$, lowers the nucleation barrier of pits such as sites 2 and 4, and induces the anisotropic dissolution morphology. The appearance of the bright region in the image (Fig. 2C) indicates the formation of a flat bottom, consistent with the energetics associated with dissolution at point defects, as predicted by the pit nucleation model (*2*).

Front velocity field analysis of the entire data sequence reveals the propagation of surface reaction fronts and their dependence on the dissolution mode (*10*). Velocities of ~5 nm/s are reached within the first few hundred seconds (Fig. 2C). Reaction fronts of the homogeneously nucleated pit are largely isotropic both in shape and velocity distributions, in agreement with the pit nucleation model. The spatial distribution of the etch pit front velocity for site 4, however, shows pronounced asymmetry and directional preference (fig. S4). The role of elastic strain in the formation of such dislocation etch pits is well established (*6*). A time-averaged and area-normalized calcite dissolution rate of 2.24 monolayers per second was calculated from the depth and lateral extent of pits in sites 3 and 3′ [fig. S3 (*10*)]. This rate is nearly 10 times faster than the separate dissolution and growth rates at equilibrium (*13*), indicating that the perturb-image procedure drives the system far from equilibrium.

The equilibrium state of the calcite surface, determined by the aqueous concentrations of carbonate species [CO_3^{2-}, HCO_3^-, $CO_{2(aq)}$, and H_2CO_3] and Ca^{2+}, is disrupted by radiolysis via photoelectrons, generating highly reactive species such as hydroxyl radicals ($OH^\bullet$) and hydrated electrons (*14*). We used a chemical kinetics model to predict

the time evolution of the above processes and their magnitudes (*10*). This model includes a network of interdependent reaction pathways including (i) radiolysis, (ii) carbonate equilibria, and (iii) calcite surface reactions (table S1). The production of radicals is described within the spur diffusion model (*9*, *15*, *16*). In the absence of radiolysis, the model correctly predicts the far–from-equilibrium calcite dissolution rate versus pH and was validated against standard geochemical reaction modeling software (figs. S6 and S7 and table S2).

Upon irradiation with the pulse sequence, the solution conditions are predicted to change through two primary mechanisms: (i) acidification of the solution by generation of H^+, with the solution pH dropping precipitously from the equilibrium initial value of 8.3 to 5.2 (Fig. 3, A and B); and (ii) scavenging of $OH^\bullet$ by the carbonate and bicarbonate ions to yield the carbonate radical CO_3^- (Fig. 3B; see fig. S8 for the time-dependent concentrations of all 25 chemical species in the system). As the acidity of the solution increases, the carbonate ion also reacts with H^+ to form bicarbonate. These factors prompt a substantial decrease of the saturation index of the solution, $\Omega(t) = \log[a_{CO_3^{2-}}(t)a_{Ca^{2+}}(t)/K_{calcite}]$, where $a(t)$ are predicted time-dependent ionic activities, and $K_{calcite}$ ($=10^{-8.48}$) is the equilibrium constant of calcite (Fig. 3D). At such a large undersaturation ($\Omega \sim -5$), $c_{carbonate}(t) << c_{eq}$, and homogeneous nucleation of pits is favored; we observed this dissolution behavior in sites 3 and 3′ (Fig. 2A). The ensuing dissolution of calcite releases Ca^{2+} and CO_3^{2-} into the solution, stabilizing the solution pH by consumption of H^+ to produce HCO_3^-, and increases the bulk concentration of dissolved carbonate (Fig. 3C). After 1700 s, the saturation index approaches its initial (equilibrium) value, and there is no longer an energetic driving force to promote further calcite dissolution. This is observed in the data, where the expansion of all dissolution modes stagnates at similar times (Fig. 2B). The average rate of calcite dissolution predicted by the kinetic model, 1.5×10^{-9} mol cm^{-2} s^{-1} or 1.8 calcite monolayers per second (Fig. 3C), is in good agreement with the observed average dissolution rate of 2.2 monolayers per second. Moreover, both of these values are in agreement with previously measured calcite dissolution rates at pH ~ 5 by other methods (*13*, *17*, *18*).

Constant irradiation led to a substantially higher degree of undersaturation (dashed lines, Fig. 3), with a predicted calcite dissolution rate that is larger by a factor of 2 (Fig. 3E). The reaction fronts of homogeneously nucleated pits (sites 1 and 2 in Fig. 4A) reached velocities of ~25 nm s^{-1} (Fig. 4B), which is five times higher than observed in the pulsed irradiation experiments (Fig. 2D). Furthermore, these conditions led to a distinct anisotropy in the spatial distribution of front velocities. Upon further time evolution, a homogeneously nucleated pit at t = 123 s (site 3), began to exhibit distortions in its reaction front (Fig. 4A). At later times (t = 240 s), the configuration of this front was reminiscent of "wormhole" instabilities (*19*). Within the first 100 s of site 3 nucleation, the distorted front propagated with a large average velocity of ~100 nm s^{-1}, eventually annihilating sites 1 and 2 by surface retreat (Fig. 4C) and covering the entire image field of view (movie S2). This mode of calcite dissolution was found to be reproducible under similar irradiation conditions and only present when the system was constantly driven by the beam probe (movies S3 to S6), corresponding to extended periods with substantial undersaturation ($\Omega < -2$). The lack of any preferred direction for the reaction fronts (Fig. 4A) and the highly anisotropic spatial distribution of the front velocities (Fig. 4B) indicate that this front undergoes an instability at these extreme undersaturations. This dissolution mode is neither controlled by preexisting strain fields of extended defects (e.g., site 4 in Fig. 2A) nor displays the isotropic front velocity field expected from pits originating at point defects (e.g., site 3 in Fig. 2A).

Instabilities are a hallmark of systems undergoing reaction-diffusion processes (*20*). Numerous

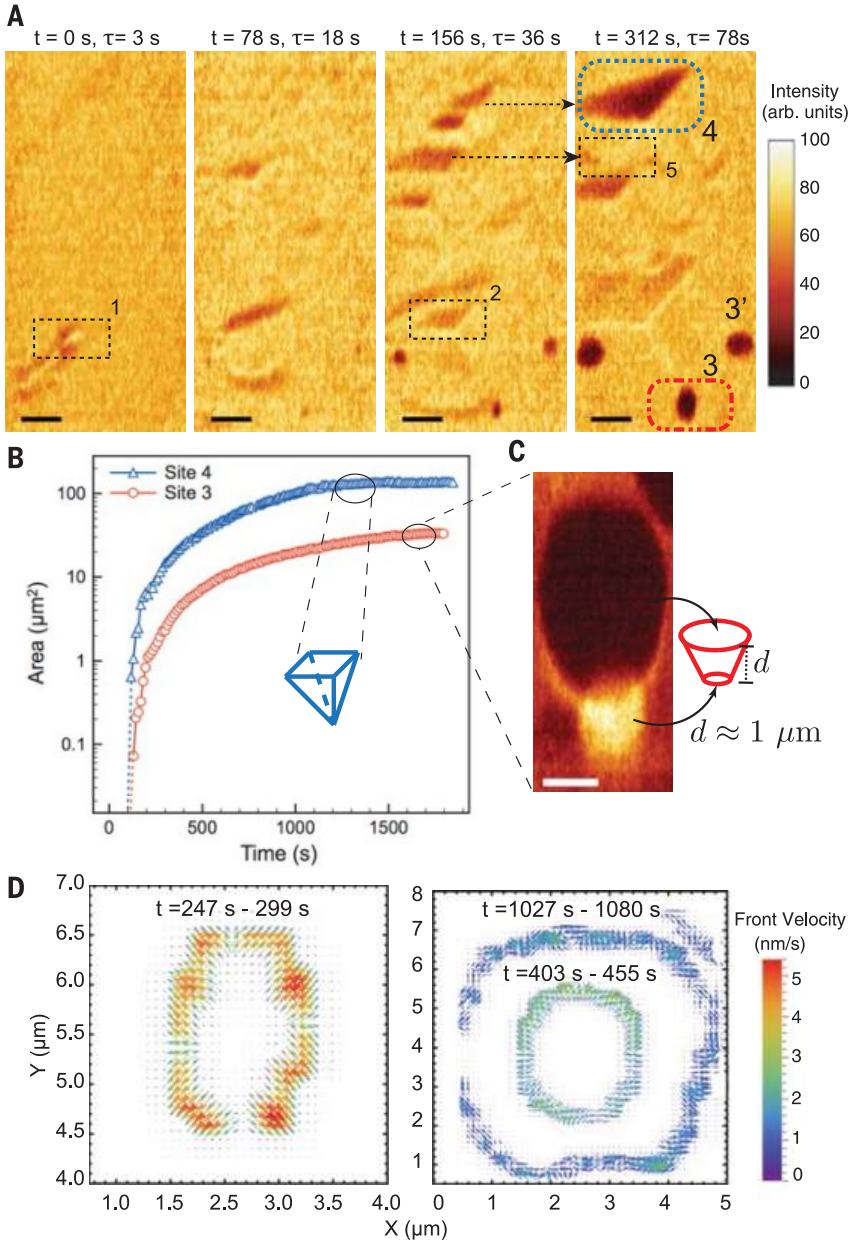


Fig. 2. Calcite dissolution driven by a pulsed x-ray beam. (**A**) X-ray images acquired during the 3-s irradiation pulse (t, experimental time; τ, total time under irradiation). Sites 1, 2, and 4, dislocation etch pit; site 3, homogeneously nucleated pit; site 5, pit annihilation by surface retreat. arb., arbitrary. Scale bars, 3 μm. (**B**) Time-dependent area measurements of sites 3 and 4. (Inset) Final shape of site 4. (**C**) The homogeneously nucleated pit of site 3 assumes the shape of a conical frustum at the indicated time. (**D**) Velocity field analysis of the reaction fronts that drive the lateral expansion of the pit in site 3. The arrows indicate the front direction, whereas their length and color give the magnitude of the front velocity in nanometers per second. The times when the reaction front vector fields were extracted are indicated in the figures.

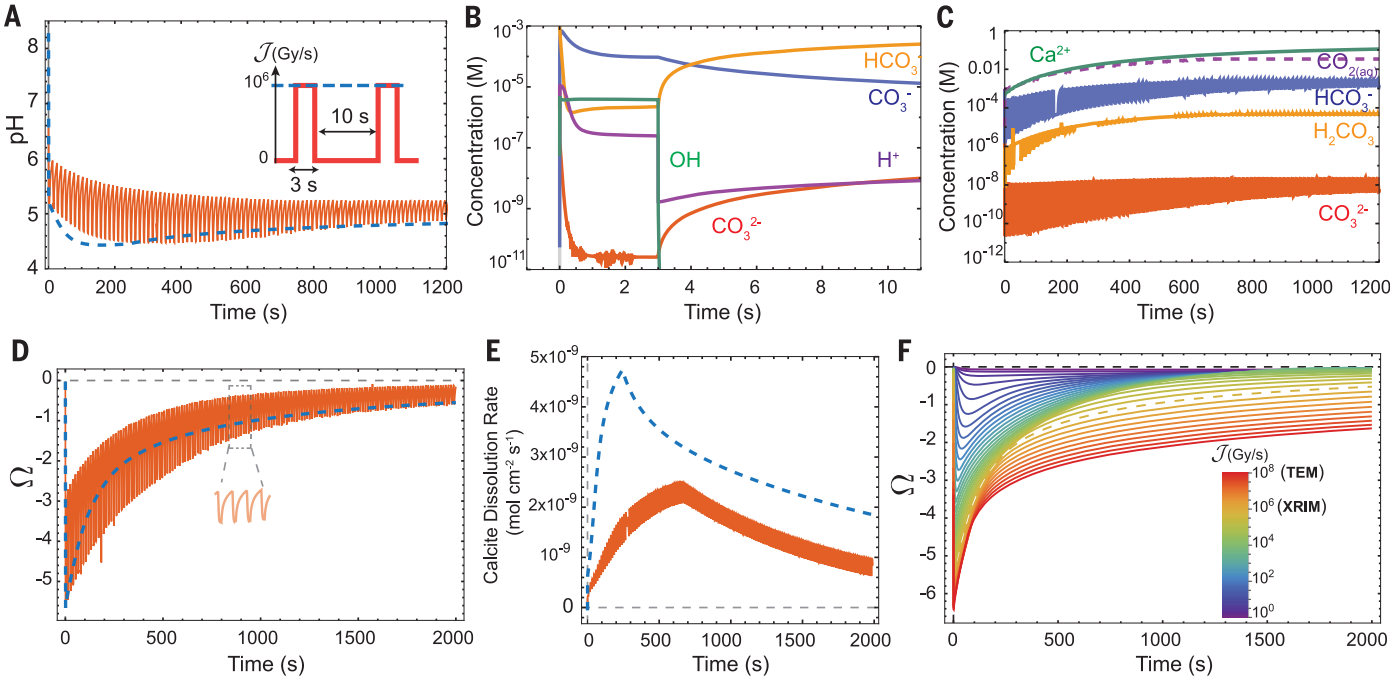


Fig. 3. The predicted evolution of the composition and saturation state of a calcite equilibrated solution due to irradiation. (**A**) Evolution of pH for continuous (dashed blue lines) and pulsed (red line) irradiation [and similarly in (D) and (E)]. (Inset) Irradiation sequences begin at t = 0 s. Dashed blue line, continuous exposure; solid red line, pulsed irradiation. (**B**) Time-dependent concentrations of solution species in response to a single x-ray square-wave pulse (off at t = 3 s). (**C**) Oscillations of solution concentrations irradiated by square-wave pulses. (**D**) Variation of the calcite solution saturation index under irradiation. (**E**) Calcite dissolution (instantaneous) rates predicted by the model (*10*). (**F**) Temporal evolution of calcite saturation index, Ω, as a function of absorbed radiation dose. XRIM subjects the system to a dose of 4 × 10^5 Gy/s under constant illumination (dashed curve) and an average of 10^5 Gy/s under pulsed illumination, whereas a typical in situ transmission electron microscope (TEM) subjects the system to ~10^8 Gy/s (*9*). At continuous doses higher than 10^4 Gy/s, the system reaches a steady state but not equilibrium.

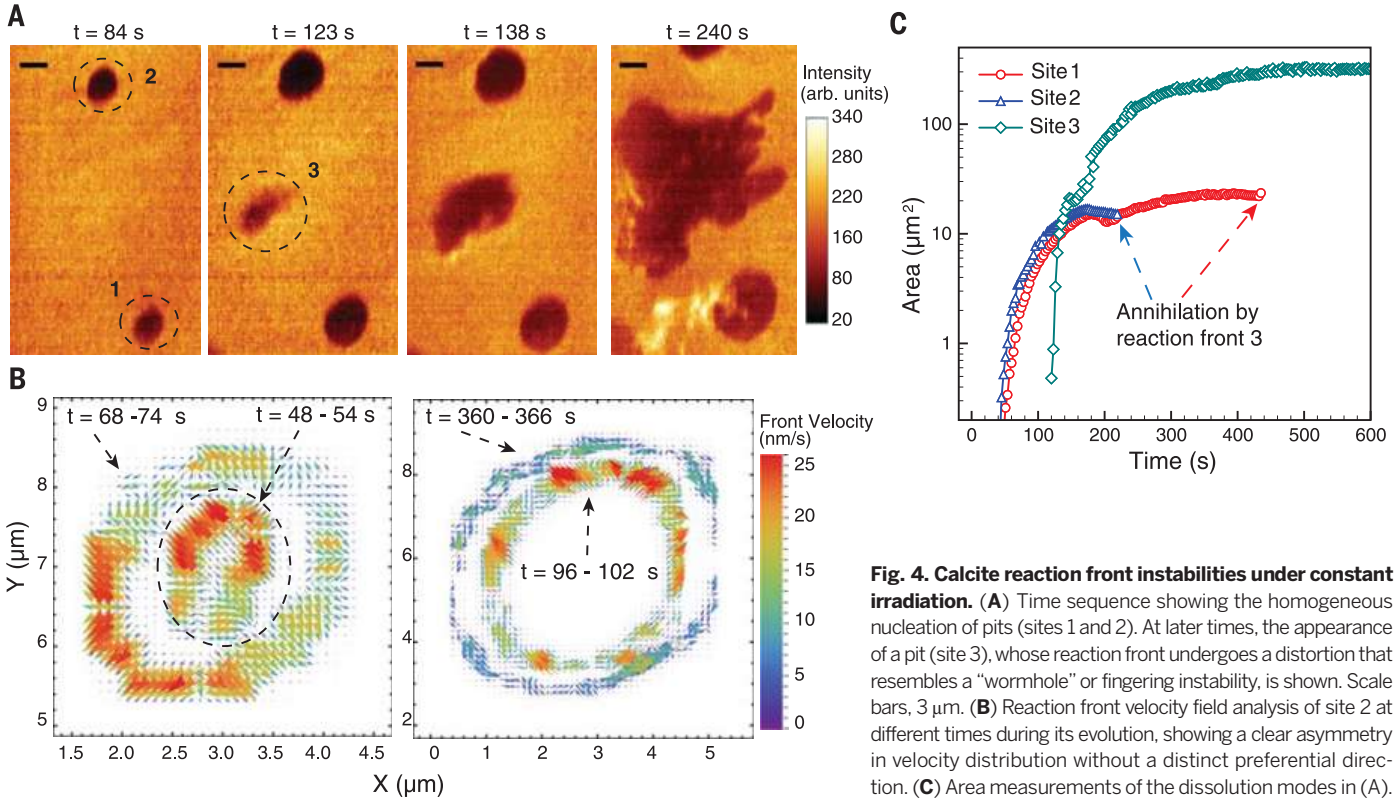


Fig. 4. Calcite reaction front instabilities under constant irradiation. (**A**) Time sequence showing the homogeneous nucleation of pits (sites 1 and 2). At later times, the appearance of a pit (site 3), whose reaction front undergoes a distortion that resembles a "wormhole" or fingering instability, is shown. Scale bars, 3 μm. (**B**) Reaction front velocity field analysis of site 2 at different times during its evolution, showing a clear asymmetry in velocity distribution without a distinct preferential direction. (**C**) Area measurements of the dissolution modes in (A).

studies, however, have established that calcite dissolution is surface-controlled (i.e., not mass transport–limited) at pH > 4 (*17*, *21*), a condition that was nominally satisfied in our experiments. The prevalence of front instabilities when the system is constantly driven by the beam probe suggests that they represent a mode of mineral/water interface reaction dynamics at conditions far from equilibrium (*19*). These instabilities therefore represent a dynamical signature of the onset of transport limitations and other dissipative processes at mineral/water interfaces.

REFERENCES AND NOTES

1. Y. Politi, T. Arad, E. Klein, S. Weiner, L. Addadi, *Science* **306**, 1161–1164 (2004).
2. S. L. Brantley, in *Kinetics of Water-Rock Interaction*, S. L. Brantley, J. D. Kubicki, A. F. White, Eds. (Springer, New York, 2008), pp. 152–210.
3. D. Gebauer, A. Völkel, H. Cölfen, *Science* **322**, 1819–1822 (2008).
4. A. F. Wallace *et al.*, *Science* **341**, 885–889 (2013).
5. A. C. Lasaga, A. Luttge, *Science* **291**, 2400–2404 (2001).
6. P. M. Dove, N. Han, J. J. De Yoreo, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15357–15362 (2005).
7. M. H. Nielsen, S. Aloni, J. J. De Yoreo, *Science* **345**, 1158–1162 (2014).
8. N. de Jonge, F. M. Ross, *Nat Nanotechnol.* **6**, 695–704 (2011).
9. N. M. Schneider *et al.*, *J. Phys. Chem. C* **118**, 22373–22382 (2014).
10. Materials and methods are available as supplementary materials on *Science* Online.
11. N. Laanait *et al.*, *J. Synchrotron Radiat.* **21**, 1252–1261 (2014).
12. P. Fenter, C. Park, Z. Zhang, S. Wang, *Nat. Phys.* **2**, 700–704 (2006).
13. L. Chou, R. M. Garrels, R. Wollast, *Chem. Geol.* **78**, 269–282 (1989).
14. E. J. Hart, *Science* **146**, 19–25 (1964).
15. A. H. Samuel, J. L. Magee, *J. Chem. Phys.* **21**, 1080 (1953).
16. J. A. Elliot, D. R. McCracken, *Fusion Eng. Des.* **13**, 21–27 (1990).
17. L. N. Plummer, T. M. L. Wigley, D. L. Parkhurst, *Am. J. Sci.* **278**, 179–216 (1978).
18. R. Shiraki, P. A. Rock, W. H. Casey, *Aquat. Geochem.* **6**, 87–108 (2000).
19. C. Steefel, in *Kinetics of Water-Rock Interaction*, S. L. Brantley, J. D. Kubicki, A. F. White, Eds. (Springer, New York, 2008), pp. 545–589.
20. I. Prigogine, G. Nicolis, *J. Chem. Phys.* **46**, 3542–3550 (1967).
21. O. S. Pokrovsky, S. V. Golubev, J. Schott, *Chem. Geol.* **217**, 239–255 (2005).

ACKNOWLEDGMENTS

This work was supported by the Geosciences Research Program of the Office of Basic Energy Sciences, U.S. Department of Energy (DOE), at Argonne National Laboratory (ANL), the University of Illinois at Chicago, and the University of Delaware. The x-ray data were collected at the Advanced Photon Source (33-ID-D), a U.S. DOE Office of Science User Facility at ANL. A portion of this research was performed by N.L. as a staff member at the Center for Nanophase Materials Sciences, a U.S DOE Office of Science User Facility at Oak Ridge National Laboratory. Primary data for this report are uncompressed video files that are available upon request from N.L and P.F. N.L. and P.F. designed the research and wrote the manuscript with input from all authors. N.L. analyzed the x-ray data and performed modeling and computations. E.B.C. and P.F. prepared the samples. All authors participated in x-ray imaging experiments.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1330/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 and S2
References (*22*–*37*)
Movies S1 to S6 (compressed video files)

10 April 2015; accepted 10 August 2015
10.1126/science.aab3272

NEURODEVELOPMENT

A mechanism for the segregation of age in mammalian neural stem cells

D. L. Moore,[1] G. A. Pilz,[1] M. J. Araúzo-Bravo,[2,3] Y. Barral,[4] S. Jessberger[1]*

Throughout life, neural stem cells (NSCs) generate neurons in the mammalian brain. Using photobleaching experiments, we found that during cell division in vitro and within the developing mouse forebrain, NSCs generate a lateral diffusion barrier in the membrane of the endoplasmic reticulum, thereby promoting asymmetric segregation of cellular components. The diffusion barrier weakens with age and in response to impairment of lamin-associated nuclear envelope constituents. Weakening of the diffusion barrier disrupts asymmetric segregation of damaged proteins, a product of aging. Damaged proteins are asymmetrically inherited by the nonstem daughter cell in embryonic and young adult NSC divisions, whereas in the older adult brain, damaged proteins are more symmetrically distributed between progeny. Thus, these data identify a mechanism of how damage that accumulates with age is asymmetrically distributed during somatic stem cell division.

Neural stem cells (NSCs) continue to give rise to new neurons throughout life in distinct areas of the mammalian brain, such as the hippocampal dentate gyrus (*1*). Adult neurogenesis results in exchange of hippocampal neurons over the human life span and has been implicated in hippocampus-dependent learning and memory (*1*, *2*). Failing or altered neurogenesis has been associated with a number of neuropsychiatric diseases such as major depression and epilepsy (*1*). The number of neurons generated is dynamically regulated. Advancing age, for example, correlates with a decline in NSC proliferation and neurogenesis, suggesting that the capacity of self-renewing NSCs to generate progeny diminishes with age (*3*, *4*). However, the cellular mechanisms governing long-term self-renewal of hippocampal NSCs and mediating the age-associated decline of their neurogenic potential remain unknown.

Budding yeast establish a lateral diffusion barrier during cell division in the membrane of the endoplasmic reticulum (ER) that functions in part to segregate senescence factors, such as extrachromosomal DNA circles and protein aggregates, asymmetrically between the mother and daughter cell, leading to similar behavior of daughter cells independent of the age (i.e., the number of cell divisions that had occurred prior) of the mother (*5*, *6*). Asymmetric distribution of senescence factors has been also described in *Drosophila* somatic stem cells, as well as in mammalian cells such as human embryonic kidney (HEK) 293T cells and human embryonic stem cells (*7*–*9*). We asked whether asymmetric segregation of senescence factors is associated with diffusion barriers in mammalian cells.

[1]Brain Research Institute, Faculty of Medicine and Science, University of Zürich, 8057 Zürich, Switzerland. [2]Biodonostia Health Research Institute, 20014 San Sebastián, Spain. [3]IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain. [4]Institute of Biochemistry, Department of Biology, ETH Zürich, 8093 Zürich, Switzerland.
***Corresponding author. E-mail: jessberger@hifo.uzh.ch**

To investigate whether rodent NSCs form a diffusion barrier in the ER membrane, we first used fluorescence loss in photobleaching (FLIP) experiments to visualize exchange of ER proteins between future daughter cells during cell division (*10*). We used green fluorescent protein (GFP)–tagged reporters of the ER lumen [using the ER-retrieval amino acid sequence Lys-Asp-Glu-Leu (KDEL); LumER-GFP] and the ER membrane (Sec61α; MemER-GFP) (*10*). NSCs expressed mCherry-tagged histone H2B to facilitate cell cycle stage identification. We performed FLIP experiments using NSCs isolated from the adult rat dentate gyrus, beginning in anaphase and continuing through the end of telophase. A small region was repetitively photobleached while the fluorescence intensity was measured in the two future daughter cells, separated by the cleavage plane (Fig. 1, A and B, and fig. S1A). Photobleaching of LumER-GFP resulted in a comparable loss of fluorescence in the bleached and unbleached daughter cells (Fig. 1, A, C, D, and F, fig. S1, A to C, and movie S1). In contrast, FLIP of MemER-GFP revealed loss of fluorescence up to the cleavage plane; the opposite daughter compartment was unaffected (Fig. 1, B, C, E, and F, fig. S1, A to C, and movie S2). The compartmentalized loss of fluorescence was not due to the distance of the photobleached area relative to the analyzed area, as bleaching regions medial to segregating chromosomes and close to the cleavage plane gave similar results (Fig. 1, G to J, fig. S1, D to I, and movies S3 and S4). Thus, NSCs establish a diffusion barrier relative to the cleavage plane that is present during cell division.

We next analyzed whether the strength of the underlying barrier is dynamically regulated with age. To confirm that in vitro aged NSCs retain some of their in vivo characteristics, we performed a 5-ethynyl-2´-deoxyuridine (EdU) pulse in hippocampal NSCs isolated from middle-aged (9-month-old) mice and young (1.5-month-old) mice. Analogous to the in vivo situation, aged NSCs in vitro were less proliferative than young

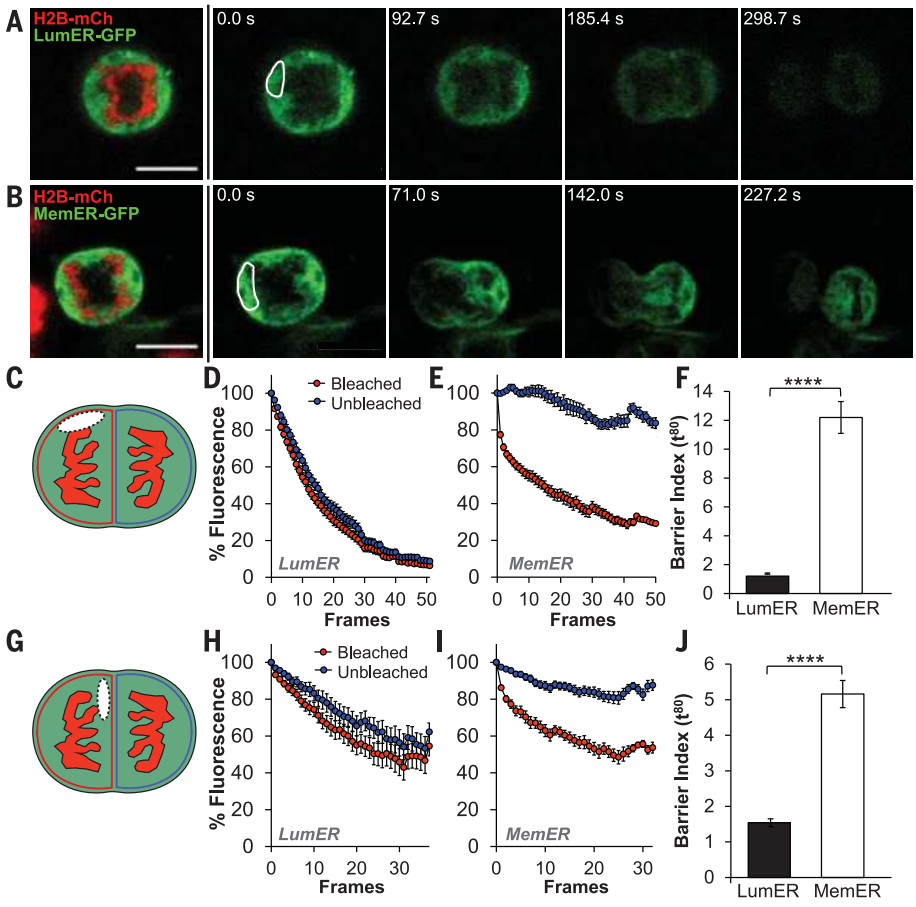


Fig. 1. Mammalian NSCs establish a lateral diffusion barrier based in the ER membrane during mitosis. (**A**) FLIP experiments in rat NSCs overexpressing LumER-GFP (green; targeted to ER lumen) or (**B**) MemER-GFP (green; targeted to ER membrane) with histone H2B-mCherry (red; DNA). Note the compartmentalized loss of fluorescence with MemER-GFP bleaching that occurs selectively on the bleached side of the dividing NSC. White outline indicates bleached region of interest (ROI). (**C**) Schematic illustrating the ROI (dashed line) repeatedly bleached in rat NSCs during time-lapse imaging for (D) to (F). Fluorescence was measured on each side of the presumptive cleavage plane throughout division, beginning in early to mid-anaphase. (**D** and **E**) Average fluorescence intensity at each time point in bleached (red) and unbleached (blue) compartments for LumER-GFP (n = 19 cells) (D) and MemER-GFP (n = 18 cells) (E); values are mean ± SEM. (**F**) Quantification of barrier strength at time t = 80 s derived from nonlinear fitted curves from (D) and (E) for LumER-GFP (black bar) and MemER-GFP (white bar; unpaired t test; mean ± SEM). (**G**) Schematic depicting the ROI (dashed line) for inner bleaching experiments used in (H) to (J). (**H** and **I**) Quantification of fluorescence loss over time for LumER-GFP (n = 20 cells) (H) and MemER-GFP (n = 22 cells) (I) when ROI is close to the cleavage plane (mean ± SEM). Bleaching began later in anaphase for an optimal ROI medial to the chromosomes, thus leading to shorter imaging length and reduced final values of total bleaching. (**J**) Quantification of barrier strength at t = 80 s based on fitted curves for LumER-GFP (black bar) and MemER-GFP (white bar) after inner ROI bleaching (unpaired t test; mean ± SEM). Scale bars, 10 μm. ****P < 0.0001.

NSCs (fig. S2, A to C). Old NSCs retained their neurogenic potential and showed gene expression profiles comparable to those of young NSCs (fig. S2, D and E, and table S1). However, in old NSCs, fluorescence substantially decreased not only in the bleached but also in the unbleached compartment upon continuous FLIP photobleaching of MemER-GFP (Fig. 2, A to E, and movies S6 and S8), indicating that the diffusion barrier weakens with age. Reduced ER compartmentalization in old NSCs was independent of potential age-associated alterations of cellular diffusion properties as measured by fluorescence recovery after photobleaching (FRAP) of MemER-GFP (fig. S2, G to J) and was also independent of anaphase-telophase duration (fig. S2F). No difference was seen between old and young NSCs in LumER-GFP FLIP experiments (fig. S2, K to O, and movies S5 and S7). Thus, the strength of the diffusion barrier decreases with age in mouse NSCs.

We next aimed to identify the molecular requirements of the mouse diffusion barrier. We focused on lamin-associated components of the nuclear envelope that become part of the ER during mitosis in mammalian cells (*11*). To disrupt lamin function in NSCs, we overexpressed a dominant-negative mutant form of lamin A called progerin that causes Hutchinson-Gilford progeria syndrome, a disease of premature aging (*12*). Retrovirus-mediated overexpression of progerin in young NSCs recapitulated the weakened diffusion barrier observed in old NSCs and led to reduced compartmentalization of fluorescence upon FLIP of MemER-GFP (Fig. 2, G, I, and J, fig. S3A, and movie S12), whereas control virus overexpression had no effect on the barrier (Fig. 2, F, H, and J, fig. S3A, and movie S10). The increased loss of fluorescence in the unbleached compartment in progerin-overexpressing cells was not due to a difference in the duration of anaphase-telophase (fig. S3B), nor to changes in the diffusion rate of MemER-GFP as measured by FRAP (fig. S3, C to F). Likewise, the dynamics of LumER-GFP were not altered by progerin overexpression (fig. S3, G to I, and movies S9 and S11). Progerin-mediated reduction in barrier strength was associated with decreased proliferation of young NSCs similar to that observed in old NSCs (fig. S3J). Thus, lamin-dependent mechanisms are required for proper barrier function and proliferation of cultured NSCs.

We next investigated whether NSCs establish a diffusion barrier when dividing within their endogenous niche. Because of the current technical limitations of imaging NSCs within their adult hippocampal niche, we switched to embryonic NSCs and electroporated embryonic day 13 (E13) mouse embryos with LumER-GFP or MemER-GFP constructs (*13*). Ex vivo slices were generated at E14, and longitudinal imaging of apical progenitors (radial glia NSCs) and basal progenitors, identified by mitotic positioning relative to the ventricular surface, began after 1 day in vitro. Because of the heterogeneity of proliferating NSCs in the developing cortex (*14*), we performed single-trace analyses of fluorescence intensity differences between the unbleached and bleached compartments for every cell upon FLIP experiments in apical and basal progenitors (fig. S4, A and B). FLIP of LumER-GFP showed no compartmentalization of fluorescence in either apical or basal progenitors (Fig. 3, A and C to F, fig. S4, A to F, and movies S15 and S18). Loss of fluorescence with MemER-GFP FLIP experiments was compartmentalized in 26% of the analyzed apical progenitors and 24% of the analyzed basal progenitors, whereas the remainder of the MemER-GFP apical and basal progenitors showed no diffusion barrier (Fig. 3, B to F, fig. S4, A to E, and movies S13, S14, S16, and S17). Thus, these findings reveal the existence of a diffusion barrier in subsets of apical progenitors and basal progenitors within their endogenous niche. Differential establishment of a diffusion barrier may contribute to the heterogeneity of apical and basal progenitors in the developing cortex (*14*).

We next asked which cellular components the barrier might segregate. In budding yeast, senescence factors such as damaged proteins and extrachromosomal DNA circles become asymmetrically segregated during division (*6*, *15*, *16*). Thus, we analyzed the distribution of ubiquitinated

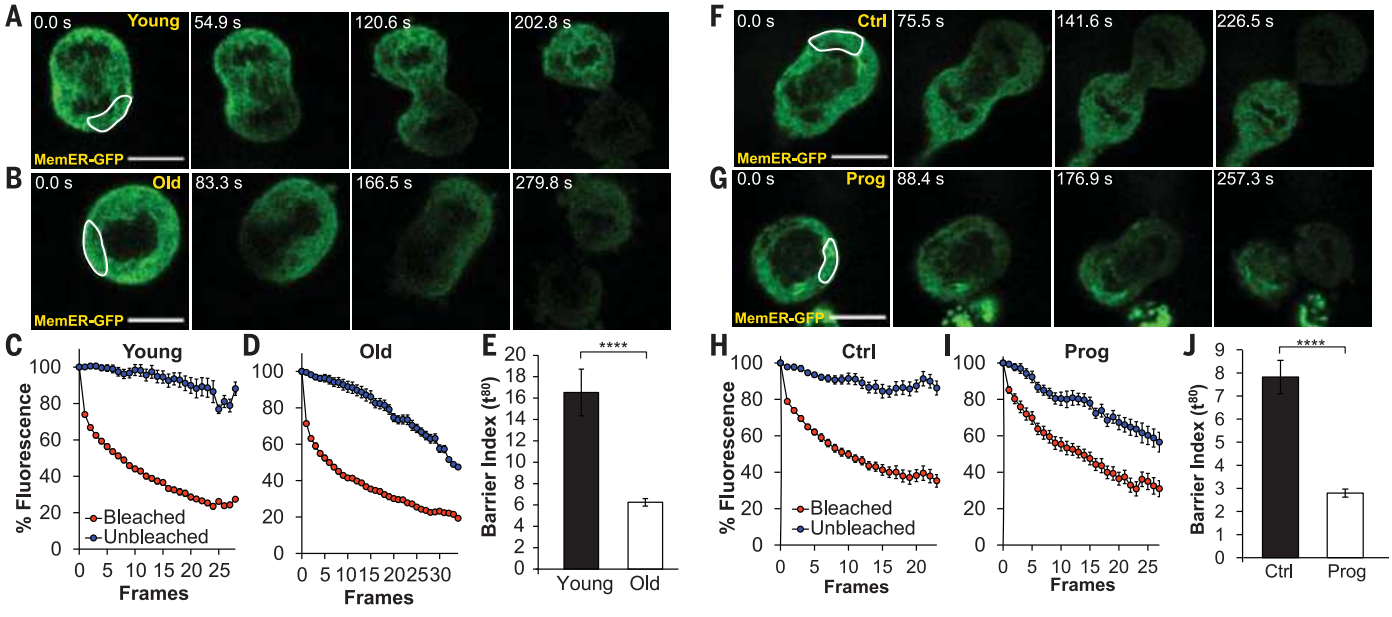


Fig. 2. The strength of the NSC diffusion barrier is weakened with age and overexpression of the lamin A mutant protein progerin. (**A** and **B**) Time course of fluorescence loss during FLIP of MemER-GFP (green) in young (A) and old (B) mouse NSCs. Note the loss of fluorescence in both compartments in old NSCs as compared to young. White outline indicates bleached ROI. (**C** and **D**) Quantification of fluorescence intensities in bleached (red) and unbleached (blue) compartments for MemER-GFP in young NSCs (n = 35 cells) (C) and old NSCs (n = 35 cells) (D); values are mean ± SEM. (**E**) Quantification of barrier strength of MemER-GFP at t = 80 s derived from nonlinear fitted curves from (C) and (D) (unpaired t test; mean ± SEM). (**F** and **G**) Representative images of a time course of fluorescence loss during FLIP of MemER-GFP (green) in young NSCs transduced with control virus (F) or progerin (G). White outline indicates bleached ROI. (**H** and **I**) Quantification of fluorescence intensity in bleached (red) and unbleached (blue) compartments for MemER-GFP in young NSCs transduced with control virus (n = 18 cells) (H) and progerin (n = 14 cells) (I); values are mean ± SEM. Progerin overexpression results in reduced compartmentalization of fluorescence loss. (**J**) Quantification of the barrier strength of MemER-GFP at t = 80 s taken from nonlinear fitted curves created from (H) and (I) (unpaired t test; mean ± SEM). Scale bars, 10 μm. ****P < 0.0001.

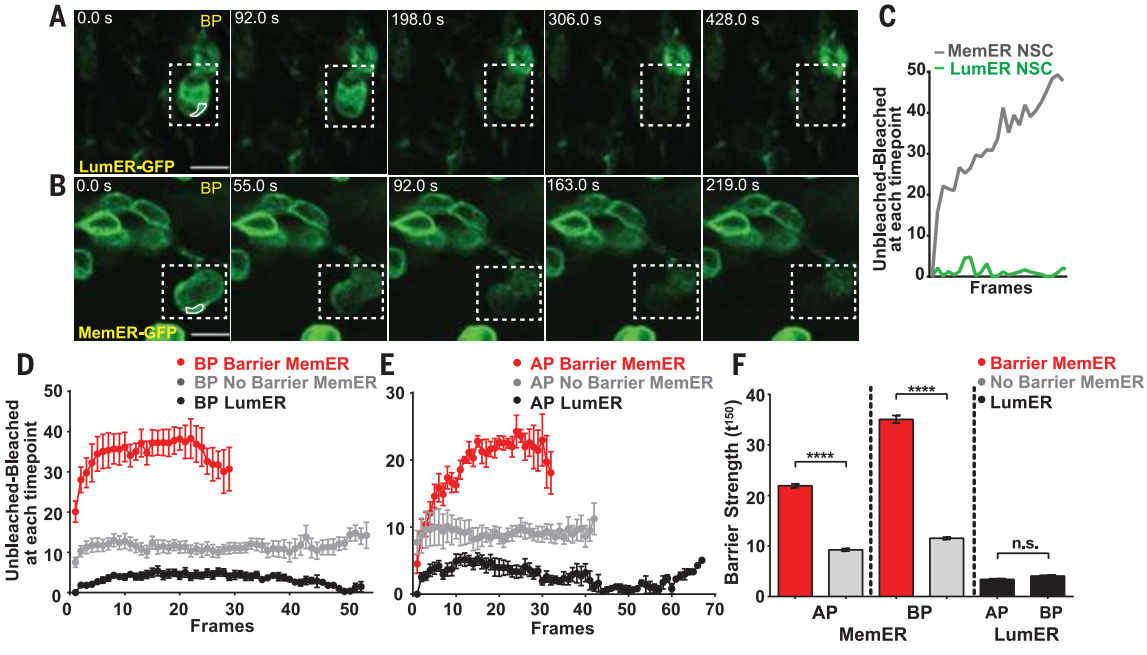


Fig. 3. Subpopulations of NSCs in the developing mouse cortex possess a lateral diffusion barrier during mitosis. (**A** and **B**) Time course of fluorescence loss during FLIP of LumER-GFP (A) and MemER-GFP (B) in basal progenitor (BP) cells of the developing cortex (E14+1 day in vitro). Note the compartmentalized loss of fluorescence with MemER-GFP bleaching selectively on the bleached side of the dividing NSC. Dashed lines indicate the cell of interest; white outline indicates bleached ROI. (**C**) Single-cell fluorescence traces plotting the difference in fluorescence intensity between the unbleached and bleached compartment at each time point reveal a compartmentalized loss of fluorescence upon FLIP in this MemER-GFP$^+$ BP (B) but not the LumER-GFP$^+$ BP (A). (**D** and **E**) Averaged single cell traces of the difference in fluorescence intensity of MemER-GFP between unbleached and bleached compartments during FLIP for each imaged BP (n = 21 cells) (D) and apical progenitor (AP; n = 13 cells) (E) reveal subpopulations that have barrier formation (red lines) and those with no barrier formation (gray lines) relative to LumER-GFP FLIP traces (black lines; n = 9 cells AP, n = 12 cells BP). (**F**) Quantification of barrier strength at t = 150 s derived from one-phase association fitted curves created from (D) and (E) (unpaired t test; mean ± SEM). Scale bars, 10 μm. ****P < 0.0001.

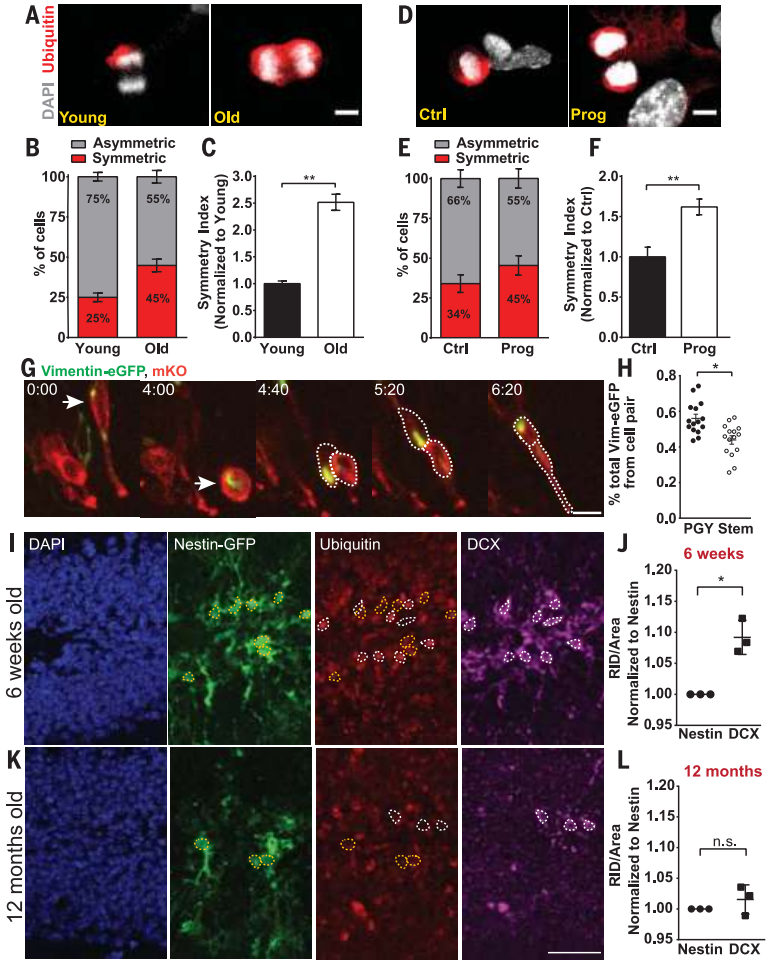


Fig. 4. Diffusion barrier strength is associated with the asymmetric segregation of damaged proteins that determine the cellular behavior of NSCs. (**A**) Ubiquitinated proteins (red) are asymmetrically distributed in young dividing NSCs (left) as compared to old NSCs (right). Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; gray). (**B**) Quantifications of symmetric (red) versus asymmetric (gray) distribution of ubiquitinated proteins in young (left) and old (right) NSCs (mean ± SEM). (**C**) A symmetry index ratio (see supplementary materials) revealed an increase in the number of old NSCs (white bar) with symmetric distribution of ubiquitinated proteins as compared to young NSCs (black bar) (unpaired *t* test; mean ± SEM). (**D**) Overexpression of progerin (right) led to a more symmetric distribution of ubiquitinated proteins relative to control virus (left). Nuclei were counterstained with DAPI (gray). (**E**) Quantifications of the symmetric (red) or asymmetric (gray) distribution of ubiquitinated proteins after control virus or progerin overexpression (mean ± SEM). (**F**) The symmetry index ratio normalized to control virus (black bar) revealed that progerin overexpression (white bar) increases the symmetric distribution of ubiquitinated proteins similarly to that of old NSCs (unpaired *t* test; mean ± SEM). Error bars on controls (Young or Ctrl) represent SEM of the original non-normalized values. (**G**) Apical progenitors overexpressing vimentin-eGFP to visualize localization of ubiquitinated proteins, and mKusabira Orange to visualize the cell outlines, were imaged in E15 mouse brain slices. The dividing NSC (arrow) and daughter cells (dashed lines) are indicated. (**H**) The intensity of vimentin-eGFP in each daughter cell was divided by the total amount of vimentin-eGFP between the two cells. The nonstem progeny (PGY, black circles) received more vimentin-eGFP at each division than did the stem daughter cell (cell that retained the apical process; empty circles) (n = 15 cell pairs; paired t test; mean ± SEM). (**I** and **K**) Nestin-GFP brains were stained against DAPI (nuclei; blue), GFP (stem cells; green), ubiquitinated proteins (Ub; red), and doublecortin (DCX; immature neurons; magenta) and imaged in 6-week-old (I) or 12-month-old (K) male mice. Outlines of confirmed cell bodies of nestin$^+$ cells (yellow outlines) and DCX$^+$ cells (white outlines) were overlaid onto the Ub channel to measure Ub intensities. (**J** and **L**) The raw integrated density (RID) of Ub was measured for each nestin-GFP$^+$ or DCX$^+$ cell body and normalized to its area before also normalizing to the background. Dots represent the averages from each animal that were normalized with nestin as 100%. In 6-week-old animals (J), DCX$^+$ cells had a significantly greater amount of Ub than in nestin$^+$ cells, whereas in 12-month-old mice (L), there was no statistical difference (P = 0.3787) between the two types of cells (n = 3 animals each; paired t test; mean ± SEM). Scale bars, 5 μm [(A) and (D)], 10 μm (G), 40 μm [(I) and (K)]. *P < 0.05, **P < 0.01.

proteins in dividing NSCs. As in mammalian stem cell populations in vitro and *D. melanogaster* embryos (*7*, *8*), young mouse NSCs displayed asymmetric segregation of ubiquitin (indicative of protein damage) in late stages of mitosis (Fig. 4, A to C). This asymmetry was reduced in old NSCs (Fig. 4, A to C). Overexpression of progerin phenocopied the effect of aging and led to a more symmetric segregation of ubiquitinated proteins (Fig. 4, D to F). Thus, a weaker barrier correlates with more symmetric distribution of damaged proteins with age.

To elucidate the functional consequences of asymmetric segregation of damaged proteins in NSCs, we established a live sensor allowing for single-cell monitoring of damaged protein segregation. Because overexpression of the sensor GFP-ubiquitin (*17*) induced cell death in NSCs, we made use of the intermediate filament protein vimentin, which asymmetrically segregates with misfolded proteins during cell division (*9*). Immunostaining of vimentin and ubiquitin in young NSCs revealed asymmetric distribution of vimentin during late mitosis. Vimentin cosegregated with ubiquitin 95.5 ± 0.8% of the time (fig. S5, A and B). Similarly, overexpressed vimentin-eGFP segregated asymmetrically in mitotic NSCs, cosegregating with endogenous ubiquitin 70.8 ± 0.9% of the time (fig. S5, C to E), and was not toxic to the cells. Thus, we asked whether asymmetric segregation of vimentin-eGFP in young dividing NSCs is associated with cell cycle length. We measured asymmetry of vimentin-eGFP at the first anaphase and calculated the length of time until the next division for each daughter cell. In 63% of dyads (P < 0.05), the daughter NSC that received more vimentin-eGFP took longer to divide than its sister cell (fig. S5F and movie S19). Thus, for rodent NSCs as for human embryonic stem cells, HEK293T cells, and certain cancer cells (*7*, *9*), ubiquitinated or damaged proteins reduce the rate of proliferation.

To analyze the asymmetric segregation of damaged proteins in situ, we electroporated E13 mouse embryos with vimentin-eGFP, performed slice cultures at E14, and began longitudinal imaging overnight. In apical progenitor divisions, the nonstem daughter cell inherited more vimentin-eGFP relative to the total amount of vimentin-eGFP, indicating that the stem cell gives away the ubiquitinated proteins to its more committed progeny (Fig. 4, G and H, and movie S20). This further supports the finding that NSCs that receive a greater amount of ubiquitinated proteins have a longer cell cycle duration, as the apical progenitor nonstem daughter cells have a lengthening of the cell cycle associated with fate restriction (*18*). To investigate damage segregation in the adult brain, we performed immunostaining on brain slices of nestin-GFP mice at 6 weeks and 12 months of age. Sections were stained with anti-doublecortin (to label newborn neurons), anti-ubiquitin (to label damaged proteins), and anti-GFP (to label hippocampal NSCs and amplify the nestin signal). Ubiquitin intensity was measured in the doublecortin$^+$ or nestin$^+$ cells. Doublecortin$^+$ cells had more ubiquitin

signal than nestin$^+$ cells in 6-week-old mice (Fig. 4, I and J). At 12 months, however, ubiquitin levels were not significantly different between NSCs and their neuronal progeny (Fig. 4, K and L) despite the general increase in ubiquitin levels in the older brain (fig. S5G). These findings establish that NSCs retain fewer damaged proteins during cell division and support the hypothesis that weakening of the diffusion barrier with age contributes to the more symmetric segregation of ubiquitinated proteins between NSCs and their neuronal progeny in vivo.

Our results show how age affects protein segregation during mammalian NSC division by altering a diffusion barrier. The diffusion barrier facilitates asymmetric segregation of damaged proteins between daughter cells, keeping the self-renewed stem cell free from damage. As in yeast, young rodent NSCs efficiently compartmentalize cellular damage, protecting the proliferative cell. Age reduces the efficiency of this compartmentalization, exposing the aged NSCs to excess cellular damage. Examining the role of the ER diffusion barrier in other mammalian cells will determine whether this represents a general mechanism for the asymmetric segregation of damage during cell divisions of somatic stem cells or other immortal cells, such as cancer cells.

REFERENCES AND NOTES

1. K. M. Christian, H. Song, G. L. Ming, *Annu. Rev. Neurosci.* **37**, 243–262 (2014).
2. K. L. Spalding *et al.*, *Cell* **153**, 1219–1227 (2013).
3. M. A. Bonaguidi *et al.*, *Cell* **145**, 1142–1155 (2011).
4. H. G. Kuhn, H. Dickinson-Anson, F. H. Gage, *J. Neurosci.* **16**, 2027–2033 (1996).
5. L. Clay *et al.*, *eLife* **3**, e01883 (2014).
6. Z. Shcheprova, S. Baldi, S. B. Frei, G. Gonnet, Y. Barral, *Nature* **454**, 728–734 (2008).
7. L. C. Fuentealba, E. Eivers, D. Geissert, V. Taelman, E. M. De Robertis, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7732–7737 (2008).
8. M. R. Bufalino, B. DeVeale, D. van der Kooy, *J. Cell Biol.* **201**, 523–530 (2013).
9. M. Ogrodnik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8049–8054 (2014).
10. C. Luedeke *et al.*, *J. Cell Biol.* **169**, 897–908 (2005).
11. C. Wandke, U. Kutay, *Cell* **152**, 1222–1225 (2013).
12. T. Dechat *et al.*, *Genes Dev.* **22**, 832–853 (2008).
13. S. C. Noctor, A. C. Flint, T. A. Weissman, R. S. Dammerman, A. R. Kriegstein, *Nature* **409**, 714–720 (2001).
14. L. M. Farkas, W. B. Huttner, *Curr. Opin. Cell Biol.* **20**, 707–715 (2008).
15. H. Aguilaniu, L. Gustafsson, M. Rigoulet, T. Nyström, *Science* **299**, 1751–1753 (2003).
16. K. A. Henderson, D. E. Gottschling, *Curr. Opin. Cell Biol.* **20**, 723–728 (2008).
17. N. P. Dantuma, T. A. Groothuis, F. A. Salomons, J. Neefjes, *J. Cell Biol.* **173**, 19–26 (2006).
18. J. T. Paridaen, W. B. Huttner, *EMBO Rep.* **15**, 351–364 (2014).

ACKNOWLEDGMENTS

Data have been deposited in NCBI's Gene Expression Omnibus with accession number GSE61367. Supplement contains additional data. We thank M. Götz for providing support for pilot slice culture experiments and conceptual input; L. Clay, B. Boettcher, A. Denoth Lippuner, X. Wang, M. Knobloch, D. Wüthrich, and R. A. Machado for experimental help; C. Balazs and M. Kirschmann for programming analyses tools; S. Rinehart for artwork; F. H. Gage, S. Aigner, and D. C. Lie for comments on the manuscript; and the light microscopy facilities of the University of Zürich Center for Microscopy and Image Analysis and ETH Zürich (ScopeM) for technical support. Supported by the Swiss National Science Foundation, the EMBO Young Investigator program, the Zürich Neuroscience Center (S.J.), the European Research Council (Y.B.), an EMBO long-term fellowship (G.A.P.), and the ETH fellow program and a Human Frontier Science Program long-term fellowship (D.L.M.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1334/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (*19*–*34*)
Data Table S1
Movies S1 to S20

9 July 2015; accepted 17 August 2015
10.1126/science.aac9868

DENGUE TYPING

Dengue viruses cluster antigenically but not as discrete serotypes

Leah C. Katzelnick,[1,2,3,4] Judith M. Fonville,[1,2,5] Gregory D. Gromowski,[3] Jose Bustos Arriaga,[3] Angela Green,[4] Sarah L. James,[1,2] Louis Lau,[4] Magelda Montoya,[4] Chunling Wang,[4] Laura A. VanBlargan,[3] Colin A. Russell,[6] Hlaing Myat Thu,[7] Theodore C. Pierson,[3] Philippe Buchy,[8] John G. Aaskov,[9,10] Jorge L. Muñoz-Jordán,[11] Nikos Vasilakis,[12,13,14] Robert V. Gibbons,[15] Robert B. Tesh,[12,13,14] Albert D.M.E. Osterhaus,[5] Ron A.M. Fouchier,[5] Anna Durbin,[16] Cameron P. Simmons,[17,18,19] Edward C. Holmes,[20] Eva Harris,[4] Stephen S. Whitehead,[3] Derek J. Smith[1,2,5]*

The four genetically divergent dengue virus (DENV) types are traditionally classified as serotypes. Antigenic and genetic differences among the DENV types influence disease outcome, vaccine-induced protection, epidemic magnitude, and viral evolution. We characterized antigenic diversity in the DENV types by antigenic maps constructed from neutralizing antibody titers obtained from African green monkeys and after human vaccination and natural infections. Genetically, geographically, and temporally, diverse DENV isolates clustered loosely by type, but we found that many are as similar antigenically to a virus of a different type as to some viruses of the same type. Primary infection antisera did not neutralize all viruses of the same DENV type any better than other types did up to 2 years after infection and did not show improved neutralization to homologous type isolates. That the canonical DENV types are not antigenically homogeneous has implications for vaccination and research on the dynamics of immunity, disease, and the evolution of DENV.

Dengue virus (DENV) infects up to 390 million people each year, and of the 96 million individuals who develop an acute systemic illness, ~500,000 experience potentially life-threatening complications, including hemorrhage and shock (*1*, *2*). The four genetic DENV types have long been thought to exist as four serotypes, and the antigenic differences between the types are believed to have a key role in the severity of disease, epidemic magnitude, viral evolution, and design of vaccines (*3*–*5*).

The description of DENV types as serotypes originated with the observation that the human immune response following primary DENV infection fully protected against challenge with viruses of the homologous type but only partially, and transiently, protected against challenge by viruses of a heterologous type (*6*). This finding was supported by in vitro neutralization experiments in which each DENV type was on average better neutralized by homologous than heterologous DENV infection antisera (*7*). The immune response immediately after a primary DENV infection varied from individual to individual, but generally was characterized by high titers of neutralizing antibodies to multiple DENV types. The neutralizing response was observed to become more DENV type–specific over time (*8*). It was later shown that antibodies to a heterologous DENV type could enhance infection in vivo and were associated with increased risk of severe disease in nature (*9*, *10*). Although antigenic variability was observed within DENV types from the earliest studies, this variation is generally considered to be substantially less than the differences between types, and not thought to modify type-specific protection (*11*, *12*). Together, the DENV types clearly form an antigenic subgroup within the genus *Flavivirus* (*13*, *14*). Analyses of envelope (E) proteins, and later full genomes, showed that the four types are as genetically divergent among themselves as sequences assigned to different viruses within the genus *Flavivirus* (*15*). These deep evolutionary divergences between DENV types were evident in the phylogenetic tree of the genetically diverse E-gene sequences of the viruses that we investigated here (Fig. 1A, fig. S1, and table S1) (*16*). Similarly, a map of amino acid differences between the E proteins revealed four compact, segregated types (Fig. 1B and fig. S2), as the number of amino acid substitutions between heterologous types far exceeded the maximum difference within a type.

However, investigations that rely on the classification of DENV into serotypes do not fully explain clinical and epidemiological phenomena.

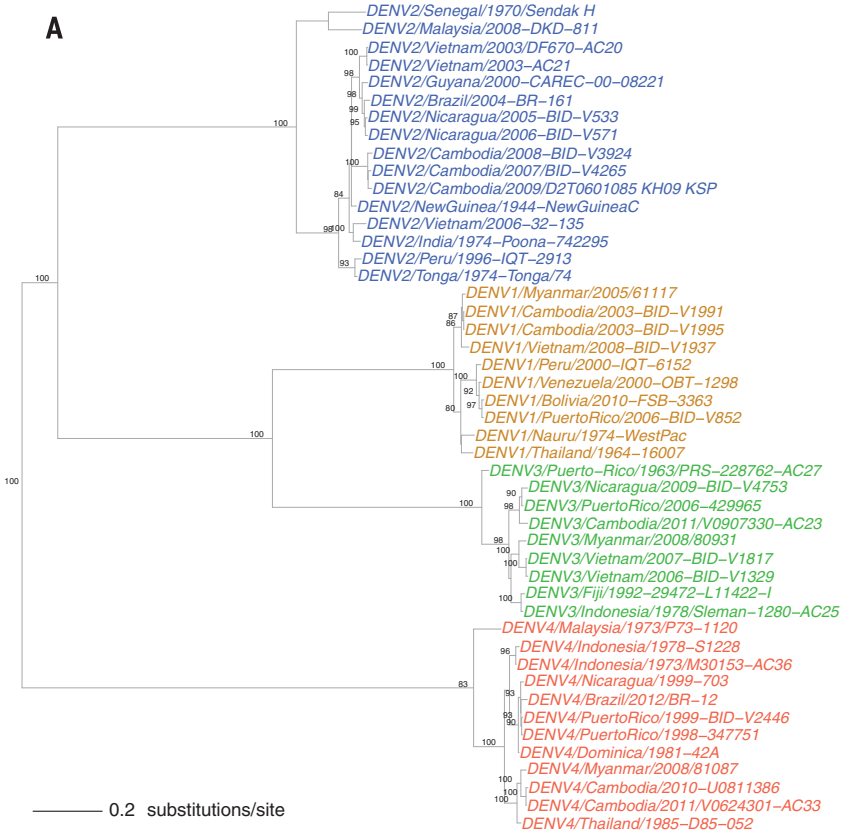


Fig. 1. Genetic analyses of the DENV panel (*n* = 47). (**A**) Phylogenetic tree showing the evolutionary relationships of DENV E gene sequences. Sequences were aligned with MAFFT, and a maximum likelihood tree was estimated using a general time-reversible model, accounting for both among-site rate variation and invariant sites (GTR+Γ_4+I). Bootstrap support values of at least 75% are shown. (**B**) Amino acid map of DENV E protein sequences (493 to 495 amino acids in length). The total amino acid differences between pairs of E sequences correspond to distances between points on the geometric display.

Nevertheless, antigenic properties are still thought to play a critical role in the biology of DENV infections. One hypothesis is that antigenic differences are critical, but that categorization by serotype alone is too coarse a measure. For example, differences in epidemic magnitude might be determined not only by the serotype but also by the antigenic differences between the particular infecting viruses that populations experience during sequential epidemics. Antigenic variation within and among the DENV types has also been hypothesized, in addition to intrinsic viral fitness and other factors, to explain phenomena including extinction and replacement of previously successful lineages and variation in disease outcome caused by genetically similar viruses (*17*–*19*). Here, we empirically test the antigenic relationships among a panel of diverse DENV isolates and reexamine the serotype concept.

Antigenic differences among viruses are caused by amino acid differences that lead to structural changes on viral proteins that modify antibody binding. The structural effect of such amino acid substitutions is difficult to predict from genetic sequences alone. In some instances substitutions have no antigenic effect; sometimes, single substitutions cause substantial antigenic change; and at other times, it takes multiple substitutions to produce an antigenic effect (*20*, *21*). Thus, today, antigenic differences must be determined by phenotype, including by an antibody neutralization assay (*13*). Most often, viruses are measured against multiple sera to form a table of neutralization data from which antigenic relationships are inferred (*22*). However, such inferences are notoriously difficult to make, and this has hindered the reliable systematic antigenic characterization of DENV. The difficulties are caused by random error, the use of diverse methods among laboratories, and the intrinsic variability among immune sera due to differences in hosts and infection histories (*23*, *24*). Moreover, neutralization data often contain apparent contradictions that are difficult to interpret, such as higher-than-homologous titers and sera that similarly neutralize multiple DENV types.

Previous antigenic analyses of DENV have addressed such challenges by using monoclonal antibodies, averaging responses of many individuals, or excluding sera with unusual patterns of reactivity. Despite careful work, these approaches have not produced a unified framework for understanding patterns across large neutralization data sets. Antigenic cartography is a method that positions viruses and antisera as points in a map, such that the distance between each virus and antiserum is derived from the corresponding neutralization titer in the tabular data. This method exploits variation in host responses to better triangulate the map, reduces the effect of some measurement errors by measuring each virus against multiple antisera

[1]Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [2]World Health Organization (WHO) Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, Cambridge CB2 3EJ, UK. [3]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [4]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA 94720-3370, USA. [5]Department of Viroscience, Erasmus MC, Rotterdam 3015 GE, Netherlands. [6]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK. [7]Department of Medical Research, Ziwaka Road, Yangon, Myanmar. [8]Institut Pasteur in Cambodia, Réseau International des Instituts Pasteur, Phnom Penh 12201, Cambodia. [9]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane 4001, Australia. [10]Australian Army Malaria Institute, Brisbane 4051, Australia. [11]Dengue Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan 00971, Puerto Rico. [12]Department of Pathology and Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, TX 77555, USA. [13]Center for Tropical Diseases, University of Texas Medical Branch, Galveston, TX 77555, USA. [14]Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX 77555, USA. [15]Department of Virology, Armed Forces Research Institute of Medical Sciences, Bangkok 10400, Thailand. [16]Center for Immunization Research, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [17]Oxford University Clinical Research Unit, Wellcome Trust Major Overseas Programme, Ho Chi Minh City, Vietnam. [18]Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7LJ, UK. [19]Department of Microbiology and Immunology, University of Melbourne, Parkville 3010, Australia. [20]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney 2006, Australia.
***Corresponding author. E-mail: djs200@cam.ac.uk**

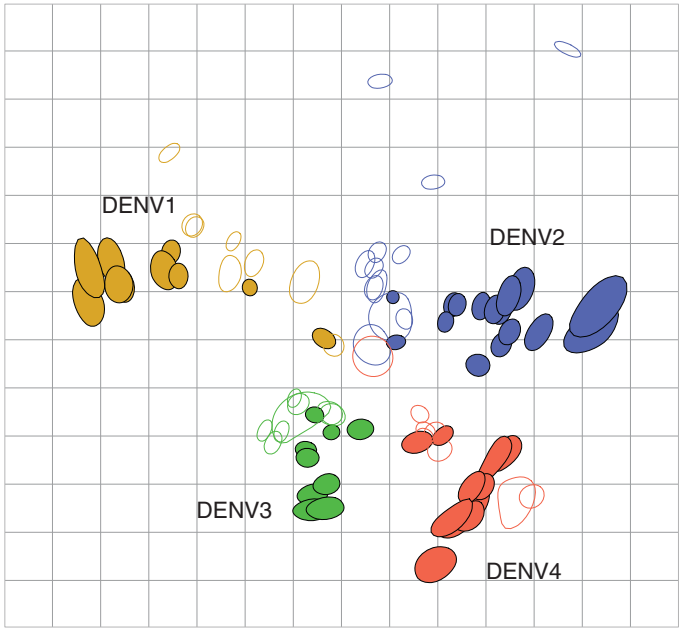


Fig. 2. Antigenic map of the DENV panel (n = 46) titrated against African green monkey antisera drawn 3 months after infection (n = 36). Each unit of antigenic distance (length of one grid-square side, measured in any direction) is equivalent to a twofold dilution in the neutralization assay. Each antiserum (open shape) and virus (closed shape) is colored according to the infecting genetic type (*16*). The size and shape of each point represent the confidence area of its position.

(and vice versa), and accurately interprets apparent contradictions in the data (*25*).

We formed the Dengue Antigenic Cartography Consortium, an open collaboration of international research laboratories, to establish empirically how DENV types relate to one another antigenically. Thirty-six African green monkeys (*Chlorocebus sabaeus*, hereafter NHP) were experimentally inoculated with diverse DENV isolates, and their sera were tested for neutralizing antibody potency against the genetically (all known genotypes), temporally (1944 to 2012), and geographically (20 countries) diverse panel of DENV isolates shown in Fig. 1 (table S1). Serum samples were taken 3 months after inoculation, and titrations were conducted using an immunofocus reduction neutralization test on mosquito cells (C6/36, *Aedes albopictus*) (tables S2 to S7 and fig. S3) (*16*, *26*). A conventional interpretation of the raw antibody neutralization titers was consistent with previous observations, both for DENV and for other flaviviruses: Antisera could generally neutralize viruses of the infecting type better than heterologous types.

The cartographic analyses fit these data with low error and were internally consistent (figs. S4, S6, and S7). Only 1% of map distances differed by more than fourfold from the measured titer (table S8). The positions of viruses and antisera were robust to different methods of calculating neutralization titers and to the exclusion of outliers (figs. S5 and S8 to S12 and table S10). Maps made with random subsets of the data set could predict excluded titers within twofold error (correlation coefficient r = 0.92 for the relation between all measured and predicted titers) (table S9).

Our analyses showed that the DENV isolates in our panel did group according to current serotype classification (Fig. 2), and the majority of viruses neighboring any given virus are of the same DENV type. However, many of the viruses were positioned as close to a virus of another DENV type as to some viruses of their own type, and the distance within and between types was comparable. Similarly, while neutralizing antisera responses clustered closely to viruses of the homologous type, almost all were at least as close to a heterologous-type isolate (tables S11 and table S12).

To examine these findings in detail, we evaluated whether the observed antigenic diversity of the virus types was also observed with human antisera and over time, and whether the neutralizing responses of individual antisera became increasingly type-specific over time.

We titrated human antisera derived from vaccination with a live-attenuated chimeric DENV vaccine against the genetically diverse DENV panel. Individuals lacking detectable neutralizing antibodies against DENV or other flaviviruses were each inoculated with one monovalent component of the National Institutes of Health DENV vaccine (n = 40 in total, 10 per DENV type). Antisera drawn 42 days after injection were titrated against the DENV panel (n = 36) by using the neutralization test on mosquito cells. The resulting antigenic map is consistent with the NHP map in that the distance between DENV types was equivalent to the spread within type, and the overall orientation of DENV1-4 was the same (Fig. 3A).

We measured the antigenic relationships among the DENV panel as recognized by antisera drawn from naturally infected individuals, who had neutralizing responses representative of the cohort study from which they were selected. Serum samples drawn from 20 Nicaraguan children in the year after their first DENV infection were titrated, by using the neutralization test on mosquito cells, against 14 viruses that captured the breadth of variation seen in the DENV panel in Fig. 2. Again, the antigenic distances among the DENV types were similar to those observed with NHP and human vaccine antisera, although the DENV4 cluster was positioned adjacent to DENV1 and DENV2 (Fig. 3B).

We also analyzed neutralization data from other studies that had used antisera from monovalent vaccine recipients and naturally infected human travelers, as well as different neutralization assays (*22*, *27*, *28*). Again, the antisera from these studies also recognized the antigenic relationships among the DENV isolates similar to the way the 3-month NHP antisera did (figs. S23 to S25).

The early antibody response is assumed to broadly neutralize all DENV types, but over time cross-type neutralization is thought to be lost so that the antibody response remaining in the months to years after infection only potently neutralizes isolates of the infecting type (*8*, *29*, *30*). We compared how antisera taken at various time points after infection recognize antigenic relationships among the DENV panel. The human antisera used to make the antigenic maps described above were taken at various times after infection, ranging from 42 days for the monovalent vaccine antisera to more than 1 year for the natural infection antisera. We also made an antigenic map of a published neutralization data set of 44 DENV isolates titrated with monkey antisera drawn 1 year after inoculation and found a similar range of antigenic variantion among the four DENV types (fig. S26) (*12*). Thus, in maps made with early (1 month) as well as late convalescent (3 months to 1 year) antisera, the antigenic relationships among diverse DENV isolates were similar to those observed with 3-month NHP antisera.

We tested if the patterns of antigenic recognition of the antisera from serially sampled individuals changed with time. We titrated antisera from the experimentally inoculated NHPs 1 month (n = 36 individuals) and 5 months (n = 16) after infection against the DENV panel. As expected, the magnitude of the neutralizing titers generally dropped between 1, 3, and 5 months (table S14). However, viruses on the 1- and 5-month antigenic maps showed the same orientation of types as the 3-month antisera. At 1 month and at 5 months after infection, 55% and 41% of the viruses, respectively, clustered as closely to a virus in a heterologous type as to some viruses of the same type (Fig. 4, A and B, and tables S11, S13, and S15). The antigenic relationships among isolates were conserved across time points (fig. S13). We thus found that the antigenic relationships among the isolates in the DENV panel were recognized similarly by early and late convalescent antisera from the same individuals.

We measured changes in neutralizing type specificity for each NHP by comparing the antiserum positions in the 1-, 3-, and 5-month antigenic maps. The antiserum positions shifted (on average, by greater than fourfold) between 1 month and 3 months, consistent with the period of somatic hypermutation and selection for affinity-matured B cells (Fig. 4A and fig. S14).

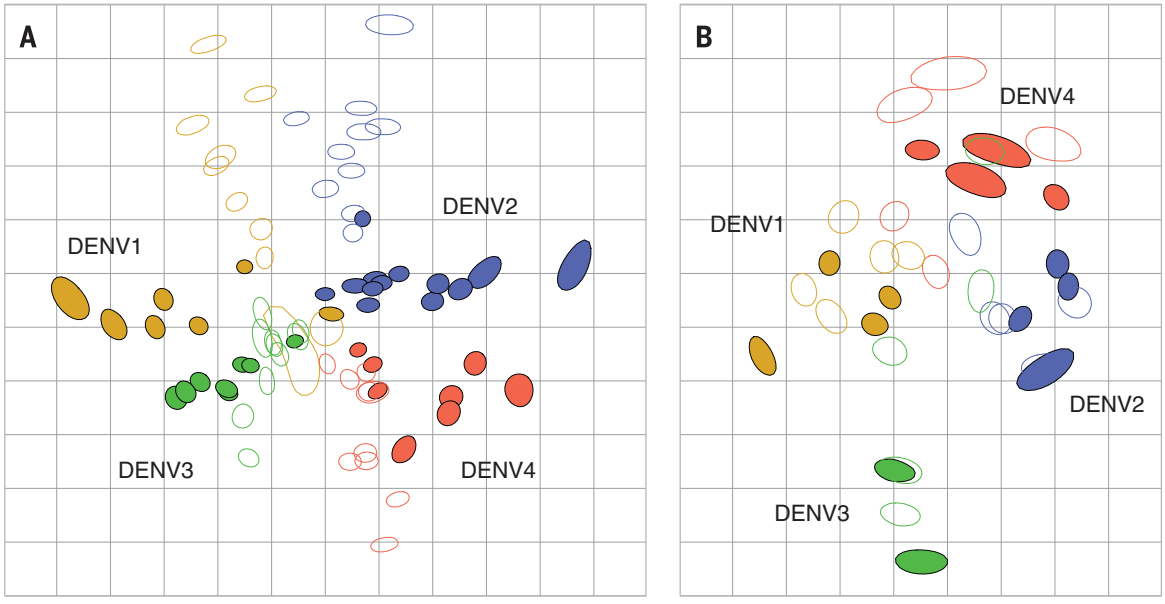


Fig. 3. Human primary infection antigenic maps. (**A**) Antisera from individuals inoculated with each monovalent component of the NIH live vaccine (10 per group) were drawn 42 days after infection and titrated against 36 viruses in the DENV panel. (**B**) Antisera from 20 Nicaraguan children drawn in the year after their first DENV infections were titrated against an antigenically diverse subset of the DENV panel (n = 14).

However, few antisera showed improved neutralization of the infecting DENV type relative to heterologous types between 1 and 3 months. The antiserum positions changed minimally between 3 and 5 months, despite a significant decline in the magnitude of titers over that period, in some cases below the assay limit of detection (Fig. 4B and table S14). Thus, we did not observe a systematic shift toward increasing neutralizing specificity to viruses of the infecting type; nor did we observe decreasing specificity toward heterotypic viruses (fig. S15 and fig. S21).

Published studies of neutralizing responses in the first year after experimental inoculation also reported stability of neutralization specificity. In one study, the ratio between homologous and heterologous neutralizing titers for 16 rhesus monkeys between 4 and 13 months after experimental inoculation was markedly consistent. NHPs that were initially type-specific remained so, and those that exhibited early cross-type titers maintained titers to those types to the end of the study period (fig. S28) (*31*). A second study following the neutralizing responses of *Aotus nancymae* monkeys for 1 to 4 months to DENV1 and DENV2 isolates showed similarly stable neutralization specificity to the infecting type and heterologous types (fig. S29) (*32*).

We further analyzed the neutralizing responses in the natural human infection data set for the type specificity of antisera obtained during the first 2 years after infection. The antisera in the map in Fig. 3B ranged in neutralizing type specificity, with 55% of antisera responses clustering as closely to a heterologous isolate as some homologous isolates. For each individual, the serum position in Fig. 3B, made with titrations conducted on mosquito cells, closely corresponded to the serum position in the map made with titrations using human cells expressing the DENV attachment factor, DC-SIGN (Fig. 3B and fig. S16). The position of the DENV4 cluster was between DENV1 and DENV2 on both maps (Fig. 3B and fig. S16). We compared the antibody titrations after 1 and 2 years for each individual and found that all maintained the pattern of neutralization, including cross-neutralization, observed in the first year after infection (figs. S17 and S18). Thus, neutralizing antibody responses in natural human DENV infections did not show a trend toward increasing type specificity even 2 years after infection.

Type-specific and cross-reactive neutralizing antibodies are thought to target distinct viral structures, and thus potentially may produce different antigenic maps (*33*). We therefore tested whether cross-reactive neutralizing antisera recognized different antigenic relationships among the DENV panel than type-specific neutralizing responses, using the serum positions of the monovalent vaccine map (Fig. 3A). Even though all 10 individuals for each DENV type were inoculated with the same vaccine component, the antisera responses to the isolates varied. Collectively, the antisera provided a coherent description of antigenic patterns among the isolates (fig. S19). The relationships among the DENV panel changed minimally between maps made with only the most central, cross-reactive 20 antisera or only the most peripheral, type-specific 20 antisera (figs. S20 and S22). Thus, the DENV type-specific and cross-reactive neutralizing responses recognized the same antigenic relationships among the DENV panel.

The antigenic characterization of any pathogen relies on the biological relevance of the assay used to generate the data. Both recent and historical studies have found statistically significant associations between neutralization titers and DENV viremia or infection outcome (*34*–*37*); however, other studies have been inconclusive (*38*, *39*). Thus, the identification of immune correlates of protection including, but not exclusively, potently neutralizing antibodies, is an active area of research for DENV (*40*–*42*). Notably, the antigenic patterns in our data are similar to those in antigenic maps that we made of DENV antibody neutralization data from other published studies using different cell lines, virus preparations, methods for detecting infected cells, and plaque or immunofocus reduction end points (figs. S23-S27) (*12*, *19*, *22*, *27*, *28*). We also found that the human antisera from natural infections titrated on mosquito cells showed neutralization profiles similar to those titrated on human cells (figs. S16 and S18). The antigenic variation we observed is thus not limited to the assay or samples that we used.

Although prior immunity to a heterologous DENV type still remains the strongest risk factor for disease, there is evidence that neutralizing responses to the particular DENV lineages circulating in a population modifies the magnitude and severity of epidemics caused by subsequent infecting lineages (*17*, *18*). In one study, cross-type neutralization provided by prior DENV1 immunity correlated with a mild epidemic caused by one lineage of DENV2, but showed no neutralization of other DENV2 lineages that in immunologically similar populations caused severe epidemics (fig. S27) (*19*). These studies and the results presented here highlight the importance of studying the specific relationship between antigenic distances as measured with neutralizing antibody titers and protection. The approach described here, in combination with global surveillance of the genetic, antigenic,

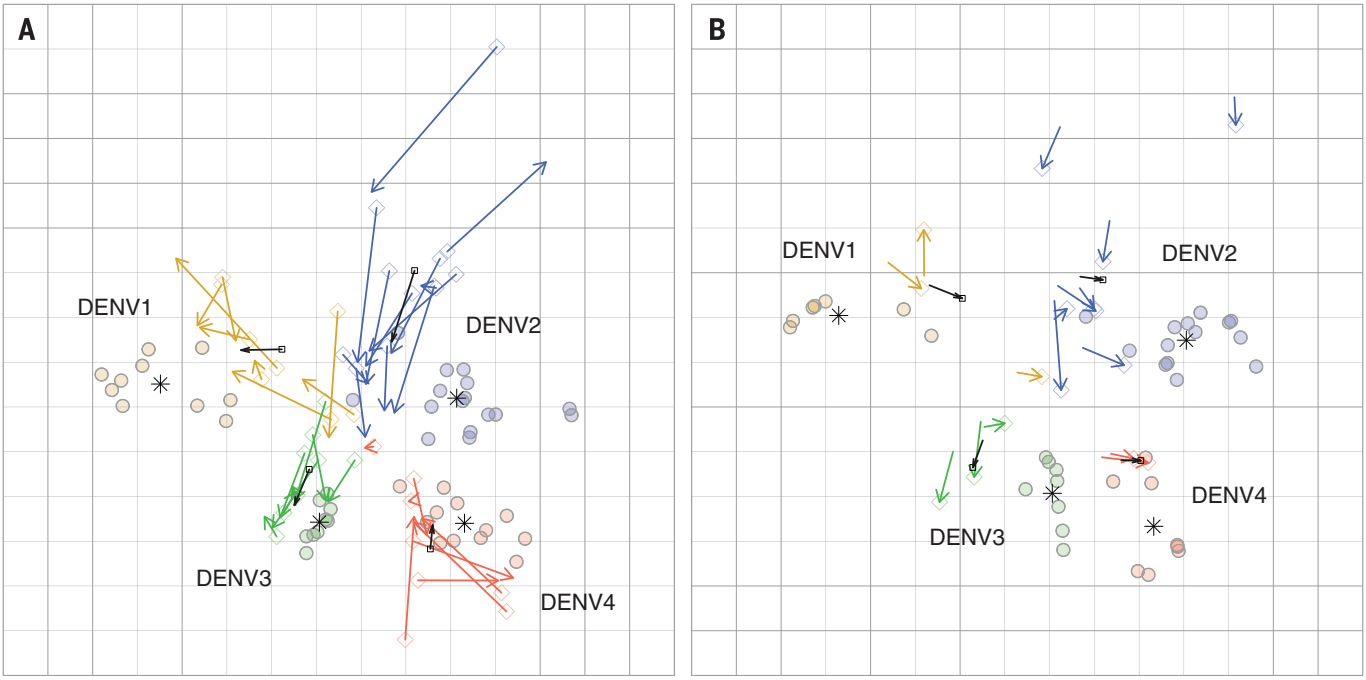


Fig. 4. Antigenic maps of the DENV panel made with antisera drawn from NHPs 1 and 5 months after infection. (**A**) An antigenic map of 47 DENV isolates titrated against 36 NHP antisera drawn 1 month after infection. Colored arrows (DENV1, yellow; DENV2, blue; DENV3, green; DENV4, red) show the change in antiserum positions between 1 and 3 months. The black arrows show the average shift in serum position for each DENV type. The star denotes the antigenic center for each DENV type. (**B**) An antigenic map of 37 DENV isolates titrated against 16 NHP antisera drawn 5 months after infection. Arrows point from positions of antisera at 3 months to the corresponding 5-month positions.

and clinical features of DENVs, as well as further detailed studies of natural infection- and vaccination-derived protection, has the potential to inform whether vaccination protects against circulating isolates as well as recognize gaps in vaccine-induced protection should they emerge over time.

The antigenic analyses shown here using 1-, 3-, and 5-month NHP antisera, human monovalent vaccine antisera, late-convalescent human natural infection antisera, and published neutralization data show that the DENV types do not fall into order as distinct serotypes. We have found that whereas DENV isolates are usually located closer to other viruses of the same type, some viruses, both modern and historical, have greater antigenic resemblance to viruses of a different type than to some viruses of the same type. We find that primary infection neutralizing antibody titers, although they drop in magnitude, do not systematically become more type-specific in the year after primary infection. As expected, individuals infected with the same or different antigens have variable patterns of neutralization, but cross-neutralizing responses consistently recognize the same antigenic relationships within the DENV panel as do the neutralizing responses that are most type-specific. These findings shift our understanding of the antigenic properties of DENV; enable more detailed study of the antigenic determinants of clinical severity, epidemic magnitude, and DENV evolution; and provide additional methods for the selection of future vaccine strains and global surveillance of the antigenic dynamics of dengue viruses.

REFERENCES AND NOTES

1. S. Bhatt *et al.*, *Nature* **496**, 504–507 (2013).
2. WHO/TDR, "Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control" (Geneva, Switzerland, 2009).
3. T. N. B. Chau *et al.*, *J. Infect. Dis.* **198**, 516–524 (2008).
4. R. S. Lanciotti, D. J. Gubler, D. W. Trent, *J. Gen. Virol.* **78**, 2279–2284 (1997).
5. C. Zhang *et al.*, *J. Virol.* **79**, 15123–15130 (2005).
6. A. B. Sabin, *Am. J. Trop. Med. Hyg.* **1**, 30–50 (1952).
7. W. M. Hammon, A. Rudnick, G. E. Sather, *Science* **131**, 1102–1103 (1960).
8. M. G. Guzman *et al.*, *Emerg. Infect. Dis.* **13**, 282–286 (2007).
9. S. B. Halstead, *J. Infect. Dis.* **140**, 527–533 (1979).
10. N. Sangkawibha *et al.*, *Am. J. Epidemiol.* **120**, 653–669 (1984).
11. M. K. Gentry, E. A. Henchal, J. M. McCown, W. E. Brandt, J. M. Dalrymple, *Am. J. Trop. Med. Hyg.* **31**, 548–555 (1982).
12. P. K. Russell, A. Nisalak, *J. Immunol.* **99**, 291–296 (1967).
13. C. H. Calisher *et al.*, *J. Gen. Virol.* **70**, 37–43 (1989).
14. K. L. Mansfield *et al.*, *J. Gen. Virol.* **92**, 2821–2829 (2011).
15. E. C. Holmes, S. S. Twiddy, *Infect. Genet. Evol.* **3**, 19–28 (2003).
16. Supplementary Materials are available on *Science* Online.
17. M. OhAinle *et al.*, *Sci. Transl. Med.* **3**, 114ra128 (2011).
18. B. Adams *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14234–14239 (2006).
19. T. J. Kochel *et al.*, *Lancet* **360**, 310–312 (2002).
20. B. F. Koel *et al.*, *Science* **342**, 976–979 (2013).
21. L. A. VanBlargan *et al.*, *PLOS Pathog.* **9**, e1003761 (2013).
22. N. Vasilakis *et al.*, *Am. J. Trop. Med. Hyg.* **79**, 128–132 (2008).
23. W. G. van Panhuis *et al.*, *J. Infect. Dis.* **202**, 1002–1010 (2010).
24. S. J. Thomas *et al.*, *Am. J. Trop. Med. Hyg.* **81**, 825–833 (2009).
25. D. J. Smith *et al.*, *Science* **305**, 371–376 (2004).
26. A. P. Durbin *et al.*, *Am. J. Trop. Med. Hyg.* **65**, 405–413 (2001).
27. A. P. Durbin *et al.*, *Virology* **439**, 34–41 (2013).
28. W. B. Messer *et al.*, *PLOS Negl. Trop. Dis.* **6**, e1486 (2012).
29. R. V. Gibbons *et al.*, *Am. J. Trop. Med. Hyg.* **77**, 910–913 (2007).
30. S. B. Halstead, G. Papaevangelou, *Am. J. Trop. Med. Hyg.* **29**, 635–637 (1980).
31. A. C. Hickey *et al.*, *Am. J. Trop. Med. Hyg.* **89**, 1043–1057 (2013).
32. T. J. Kochel *et al.*, *J. Infect. Dis.* **191**, 1000–1004 (2005).
33. W. Dejnirattisai *et al.*, *Nat. Immunol.* **16**, 785(2015).
34. D. Buddhari *et al.*, *PLOS Negl. Trop. Dis.* **8**, e3230 (2014).
35. C. A. Sariol, L. J. White, *Front. Immunol.* **5**, 452 (2014).
36. S. C. Kliks, S. Nimmanitya, A. Nisalak, D. S. Burke, *Am. J. Trop. Med. Hyg.* **38**, 411–419 (1988).
37. D. H. Libraty *et al.*, *PLOS Med.* **6**, e1000171 (2009).
38. T. P. Endy *et al.*, *J. Infect. Dis.* **189**, 990–1000 (2004).
39. A. Sabchareon *et al.*, *Lancet* **380**, 1559–1567 (2012).
40. S. A. Plotkin, *Clin. Infect. Dis.* **56**, 1458–1465 (2013).
41. S. Mukherjee *et al.*, *J. Virol.* **88**, 7210–7220 (2014).
42. G. N. Malavige, G. S. Ogg, *J. Clin. Virol.* **58**, 605–611 (2013).

ACKNOWLEDGMENTS

We express our gratitude to the members of the Dengue Antigenic Cartography Consortium, named in the supplementary materials, for their advice and contributions to the Consortium to date. We thank D. Burke, N. Lewis, E. Selkov, E. Skepner, A. Mosterín, R. Mögling, S. Wilks, T. Kotarba, and V. Duong for their technical expertise. M. Melendrez, J. Hang, R. Jarman, S. M. Cave, S. G. Widen, T. G. Wood, and V. Duong assisted with virus sequencing. C. Firestone and M. Galvez assisted with neutralization assay titrations. This research was supported in part by the Intramural Research Program of the U.S. NIH, National Institute of Allergy and Infectious Diseases (NIAID), European Union (EU) FP7 programs EMPERIE (223498) and ANTIGONE (278976), Human Frontier Science Program (HFSP) program grant P0050/2008, the NIH Director's Pioneer Award DP1-OD000490-01, the FIRST program from the Bill and Melinda Gates Foundation, and the Instituto Carlos Slim de la Salud (E.H.). The antigenic cartography toolkit was in part supported by NIAID-NIH Centers of Excellence for Influenza Research and Surveillance contracts HHSN266200700010C and

HHSN272201400008C for use on influenza virus. L.C.K. was supported by the Gates Cambridge Scholarship and the NIH Oxford Cambridge Scholars Program. J.M.F. was supported by a Medical Research Council Fellowship (MR/K021885/1) and a Junior Research Fellowship from Homerton College Cambridge. E.C.H. was supported by an National Health and Medical Research Council Australia Fellowship. N.V. and R.B.T were supported by NIH contract HHSN272201000040I/HHSN27200004/D04. The viruses and sera used in this study are covered by standard material transfer agreements at the home institutions of S.S.W., C.P.S., E.H., P.B., J.G.A., J.L.M.J., N.V., and R.B.T. A.D.M.E.O. is a professor and director of Artemis One Health Utrecht, The Netherlands; Chief Scientific Officer Viroclinics Biosciences BV, the Netherlands; and a Board Member of Protein Sciences USA. P.B. performed this work while at the Institut Pasteur in Cambodia, but since June 2015, is with GlaxoSmithKline vaccines in Singapore, and has stock options with GSK. C.P.S. is a paid consultant to GSK Pharma, GSK Vaccines, and Merck and has received a grant and consulting payments to his institution from Sanofi Pasteur. The sequences used in this study are available from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and are listed in table S1. Files used for genetic analyses are available as supplementary data files. The NIH monovalent DENV vaccines trials (ClinicalTrials.gov identifiers: NCT00473135 NCT00920517, NCT00831012, NCT00831012) were performed under an investigational new drug application reviewed by the U.S. Food and Drug Administration and approved by the Institutional Review Board at the University of Vermont and Johns Hopkins University. Informed consent was obtained in accordance federal and international regulations (21CFR50, ICHE6). The Pediatric Dengue Cohort Study in Managua, Nicaragua, was approved by the Institutional Review Boards of the Nicaraguan Ministry of Health and the University of California, Berkeley. Parents or legal guardians of all subjects provided written informed consent, and subjects 6 years of age and older provided assent.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1338/suppl/DC1
Materials and Methods
Figs. S1 to S29
Tables S1 to S15
Data Files S1 to S9
References (*43*–*66*)

6 May 2015; accepted 6 August 2015
10.1126/science.aac5017

HUMAN GENETICS

Greenlandic Inuit show genetic signatures of diet and climate adaptation

Matteo Fumagalli,[1,2]* Ida Moltke,[3]* Niels Grarup,[4] Fernando Racimo,[2] Peter Bjerregaard,[5,6] Marit E. Jørgensen,[5,7] Thorfinn S. Korneliussen,[8] Pascale Gerbault,[1,9] Line Skotte,[3] Allan Linneberg,[10,11,12] Cramer Christensen,[13] Ivan Brandslund,[14,15] Torben Jørgensen,[10,16,17] Emilia Huerta-Sánchez,[18] Erik B. Schmidt,[17,19] Oluf Pedersen,[4] Torben Hansen,[4]† Anders Albrechtsen,[3]† Rasmus Nielsen[2,20]†

The indigenous people of Greenland, the Inuit, have lived for a long time in the extreme conditions of the Arctic, including low annual temperatures, and with a specialized diet rich in protein and fatty acids, particularly omega-3 polyunsaturated fatty acids (PUFAs). A scan of Inuit genomes for signatures of adaptation revealed signals at several loci, with the strongest signal located in a cluster of fatty acid desaturases that determine PUFA levels. The selected alleles are associated with multiple metabolic and anthropometric phenotypes and have large effect sizes for weight and height, with the effect on height replicated in Europeans. By analyzing membrane lipids, we found that the selected alleles modulate fatty acid composition, which may affect the regulation of growth hormones. Thus, the Inuit have genetic and physiological adaptations to a diet rich in PUFAs.

Previous studies have attempted to understand the genetic basis of human adaptation to local environments, including cold climates and a lipid-rich diet (*1*). A recent study found evidence that a coding variant in *CPT1A*, a gene involved in the regulation of long-chain fatty acid, has been the target of strong positive selection in native Siberians, possibly driven by adaptation to a cold climate or to a high-fat diet (*2*). Another study found evidence that adaptation to the traditional hypoglycemic diet of Greenlandic Inuit may have favored a mutation in *TBC1D4* that affects glucose uptake and occurs at high frequency only among the Inuit (*3*). However, knowledge about the genetic basis of human adaptation to cold climates and lipid-rich diets remains limited.

Motivated by this, we performed a scan for signatures of genetic adaptation in the population of Greenland. The Inuit ancestors of this population arrived in Greenland less than 1000 years ago (*4*), but they lived in the Arctic for thousands of years before that (*5*). As such, they have probably adapted to the cold Arctic climate and to their traditional diet, which has a high content of omega-3 polyunsaturated fatty acids (PUFAs) derived from seafood (*6*) and a content of omega-6 PUFAs that is lower than in Danish controls (*7*).

We analyzed data from previously genotyped Greenlandic individuals (*3*) by using the Illumina MetaboChip (*8*), which is an array enriched with single-nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWASs) associated with cardiometabolic phenotypes. As a result of recent admixture, modern Greenlanders have, on average, 25% genetic European ancestry (*9*). To get a representative sample of the indigenous Greenlandic Inuit (GI), we analyzed the subset of 191 individuals that had less than 5% estimated European ancestry per individual (0.5% on average) (*9*). We combined the data from these individuals with the MetaboChip data from 60 individuals of European ancestry (CEU) and 44 Han Chinese individuals (CHB) from the HapMap Consortium (fig. S1) (*10*).

To detect signals of positive selection, we used the population branch statistic (PBS) (*11*), which identifies alleles that have experienced strong changes in frequency in one population (GI) relative to two reference populations (CEU and CHB) (*5*). A sliding window analysis identified several SNP windows with high PBS values, indicative of selection (Fig. 1 and table S1).

The strongest signal of selection is located within a region on chromosome 11 (Fig. 1A) and encompasses five genes: two open reading frames, C11orf10 (*TMEM258*) and C11orf9 (*MYRF*); and three fatty acid desaturases, *FADS1*, *FADS2*, and *FADS3*. The SNP with the highest PBS value falls within *FADS2*. The function of *FADS3* is not known; *FADS1* and *FADS2* encode delta-5 and delta-6 desaturases, which are the rate-limiting steps in the conversion of linoleic acid (omega-6) and α-linolenic acid (omega-3) to the longer, more unsaturated and biologically active eicosapentaenoic acid (EPA, omega-3), docosahexaenoic acid (DHA, omega-3), and arachidonic acid (omega-6).

[1]Department of Genetics, Evolution, and Environment, University College London, London WC1E 6BT, UK. [2]Department of Integrative Biology, University of California–Berkeley, Berkeley, CA 94720, USA. [3]The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark. [4]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark. [5]National Institute of Public Health, University of Southern Denmark, 1353 Copenhagen, Denmark. [6]Greenland Center for Health Research, University of Greenland, Nuuk, Greenland. [7]Steno Diabetes Center, 2820 Gentofte, Denmark. [8]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark. [9]Department of Anthropology, University College London, London WC1H 0BW, UK. [10]Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark. [11]Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark. [12]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [13]Department of Medicine, Lillebaelt Hospital, Vejle, Denmark. [14]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark. [15]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [16]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [17]Faculty of Medicine, University of Aalborg, Aalborg, Denmark. [18]School of Natural Sciences, University of California–Merced, Merced, CA 95343, USA. [19]Department of Cardiology, Aalborg University Hospital, 9100 Aalborg, Denmark. [20]Department of Statistics, University of California–Berkeley, Berkeley, CA 94720, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: torben.hansen@sund.ku.dk (T.H.); albrecht@binf.ku.dk (A.A.); rasmus_nielsen@berkeley.edu (R.N.)**

Polymorphisms in *FADS1* and *FADS2* are associated with increased levels of plasma and erythrocyte delta-5 desaturases in Alaskan Inuit (*12*) as well as with levels of PUFA in blood and breast milk (*13*, *14*).

We also found signals of selection in a region on chromosome 1 (Fig. 1A), which encompasses *WARS2*, a mitochondrial tryptophanyl-tRNA synthetase, and *TBX15*, a transcription factor member of the T-box family. Within this region, the SNP with the highest PBS value is located upstream of *WARS2*. Polymorphisms in or near *WARS2* and *TBX15* have been shown to be associated with numerous phenotypes among individuals of European descent, including waist-hip ratio (*15*). Based on linkage disequilibrium (LD) patterns in Greenlandic Inuit, the results from (*15*) suggest that the allele that occurs frequently in Greenlandic Inuit may decrease the waist-hip ratio. *TBX15* plays a role in the differentiation of brown (subcutaneous) and brite (typically inguinal) adipocytes (*16*). The latter, upon stimulation by exposure to cold, can differentiate into cells capable of expressing UCP1 (uncoupling protein 1), which produces heat by lipid oxidation. Therefore, *TBX15* may be associated with adaptation to cold in Inuit.

FN3KRP shows evidence of selection as well (Fig. 1A). *FN3KRP* encodes an enzyme that catalyzes fructosamines, psicosamines, and ribulosamines. This protein protects against nonenzymatic glycation, an oxidative process that is associated with various pathophysiologies (*17*). A high intake of PUFAs is associated with increased oxidative stress (*18*); it is possible that the alleles affected by selection in *FN3KRP* counteract the negative fitness caused by a PUFA-rich diet. A list of additional candidate regions under positive selection is presented in tables S2 and S3.

To corroborate our results from the SNP chip–based analysis described above, we also calculated PBS values (table S4) for exome sequencing data from 18 unrelated GI individuals (*3*), combined with data from 85 CEU individuals and 97 CHB individuals from the 1000 Genomes Project (fig. S1) (*19*).

These analyses identified two high-scoring genes (table S5): *DSP*, a gene associated with cardiomyopathy (*20*), and *ANGPTL6*, a gene that counteracts high-fat diet–induced obesity and related insulin resistance through increased energy expenditure (*21*). Gene ontology enrichment analyses of genes under selection revealed enriched muscle- and heart-development categories, similar to those positively selected in polar bears (table S6) (*5*, *22*).

In addition, these analyses reproduced the strong signal observed in the *FADS1-FADS2-FADS3* region, even though the SNPs with the highest PBS values are not detected by the system used for exome capture (Agilent SureSelect; fig. S2), and this region has the SNP with the strongest signal of selection (i.e., highest PBS value) in any of the data analyzed. We therefore focused on this region for the rest of this study. On the basis of an inferred demographic model (*5*), we estimated a divergence time between CHB and GI of 23,250 years before the present (yr B.P.), unidirectional gene flow from GI to CHB at some point in the history of these populations, and a reduced effective population size of GI (effective population size = 1550). The estimated model (fig. S3A) fits the observed joint site frequency spectrum (fig. S4), and the PBS value for the *FADS* region is a strong outlier, corroborating the idea that selection probably has affected this region (fig. S5).

Using an approximate Bayesian computation approach, we also estimated the starting time and intensity of selection, *s* (*5*). Because of the high LD within the region and the fact that our data were from SNP chip (fig. S6), we could not pinpoint the causative SNP(s) by means of population genetic analyses; we therefore used the SNP with the highest PBS value (reference SNP identification number rs74771917) as a proxy. This SNP has a derived allele frequency of 0.98 in GI, 0.025 in CEU, and 0.16 in CHB. Our analyses produced maximum a posteriori probability (MAP) estimates of the selection starting time, 19,751 yr B.P. [95% Bayesian credible interval (BCI): 2499 to 22,771 yr B.P.] (figs. S3B and S7), and of *s*, 3.13% (95% BCI: 0.98 to 19.49%) (fig. S3C). These results suggest that selection began to act on these genes long before the earliest settlement of Inuit in Greenland (*4*). In population samples from the HGDP-CEPH (Human Genome Diversity Project–Centre d'Etude du Polymorphisme Humain) database, the selected allele of rs74771917 has much higher frequencies among Native Americans than it does among East Asians (fig. S8) (*23*), suggesting that selection began to act before the Inuit split from the Native Americans, when their common ancestors lived in or around Beringia (*24*).

Six SNPs in the *FADS* region (Table 1) have PBS values above 2, suggesting that they have been subjected to strong selection. One of these SNPs, rs174570, is associated with circulating high-density lipoprotein, low-density lipoprotein (LDL), and total cholesterol levels in Europeans (*25*). We therefore tested for associations between the top six SNPs and 13 metabolic and anthropometric phenotypes in Greenlanders by analyzing data from the Greenlandic cohorts IHIT (Inuit Health in Transition) and B99 (Greenland Population Study 1999), which include 2733 and 1331 genotyped individuals, respectively (*3*). We analyzed the cohorts separately, combined the results in a meta-analysis (*5*), and found marginally significant associations with multiple phenotypes, including body-mass index, fasting serum insulin, and fasting serum LDL cholesterol (tables S7 to S12). In all cases, the derived (selected)

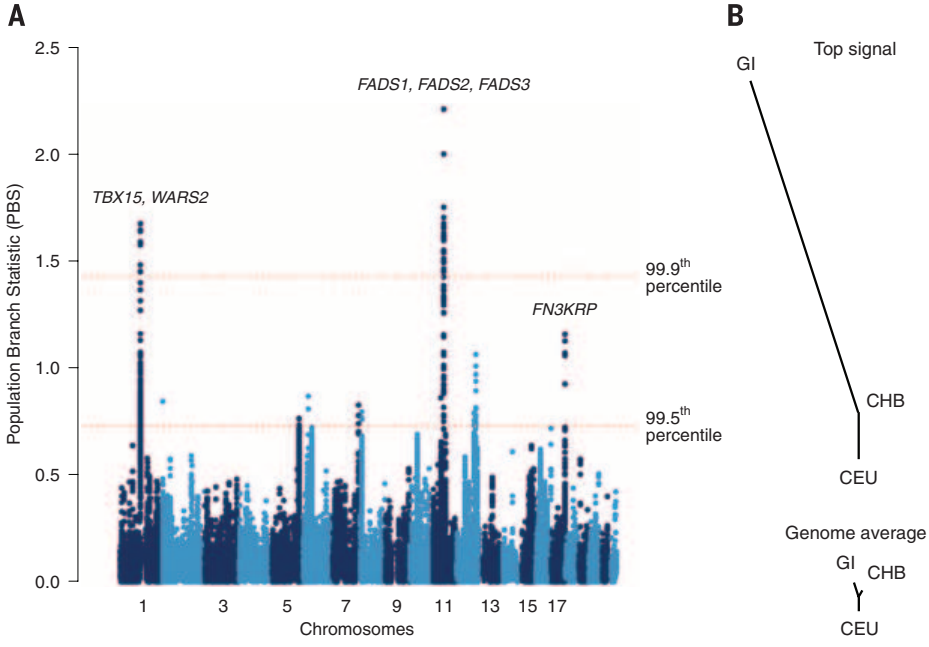


Fig. 1. Results from a genome-wide scan for positive selection. (**A**) PBS values in windows of 20 SNPs, using a step size of 5 SNPs. The 99.5th and 99.9th percentiles of the empirical distribution are shown as red dashed horizontal lines. Names of genes associated with the highest peaks are shown. (**B**) Evolutionary trees underlying the strongest signal of selection. The bottom panel shows genomic-average branch lengths based on F_{ST} (fixation index, a measure of population genetic differentiation) for GI, CEU and CHB branches (bottom); the top panel shows branch lengths for the SNPs in the window with the highest PBS values, indicating substantial changes in allele frequencies along the GI branch.

allele was associated with a reduction in the phenotypic value. The strongest association was with body weight ($P = 1.1 \times 10^{-6}$; rs7115739) and height ($P = 0.00012$; rs7115739) (table S10). Both of these associations remained significant after Bonferroni correction for testing for association between 13 phenotypes and six SNPs. To further validate the association with height, we genotyped an additional Greenlandic cohort, known as BBH, consisting of 541 Greenlandic individuals who live in Denmark and for whom height information is available. When we added these data to the meta-analysis of height, the association signal for rs7115739 became even stronger ($P = 4.6 \times 10^{-7}$). Moreover, the per-allele effect size estimates for the derived allele for height and weight are −0.66 cm and −2.2 kg in IHIT and −1.2 cm and −2.4 kg in B99 (Fig. 2, A and B, and table S10). As mentioned, the statistical method that we used accounts for admixture. Furthermore, we observed an effect both in Greenlanders with little or no European ancestry and in Greenlanders with more than 40% European ancestry when we stratified the data on the basis of ancestry proportions, which we would not expect if the association signal was caused by admixture in our data (fig. S9). These observations indicate that our association results are not caused by insufficient correction for admixture.

The six SNPs with the highest PBS values are also polymorphic in Europeans (Table 1). However,

Table 1. Annotation for the top six SNPs under positive selection in Greenlandic Inuit. DAFs for each population (CEU, CHB, and GI) and PBS values are reported, along with the genomic position for each SNP.

Position*	Reference SNP identification number	Alleles†	DAF			PBS
			CEU	CHB	GI	
chr11:61627960	rs74771917	C/T	0.025	0.16	0.98	2.67
chr11:61631510	rs3168072	A/T	0.017	0.18	0.98	2.64
chr11:61632310	rs12577276	A/G	0.017	0.18	0.98	2.64
chr11:61641717	rs7115739	G/T	0.017	0.22	0.98	2.54
chr11:61624414	rs174602	C/T	0.80	0.73	0.01	2.11
chr11:61597212	rs174570	C/T	0.16	0.34	0.99	2.056

*Positions refer to human genome assembly hg19. †Alleles are coded as ancestral/derived states.

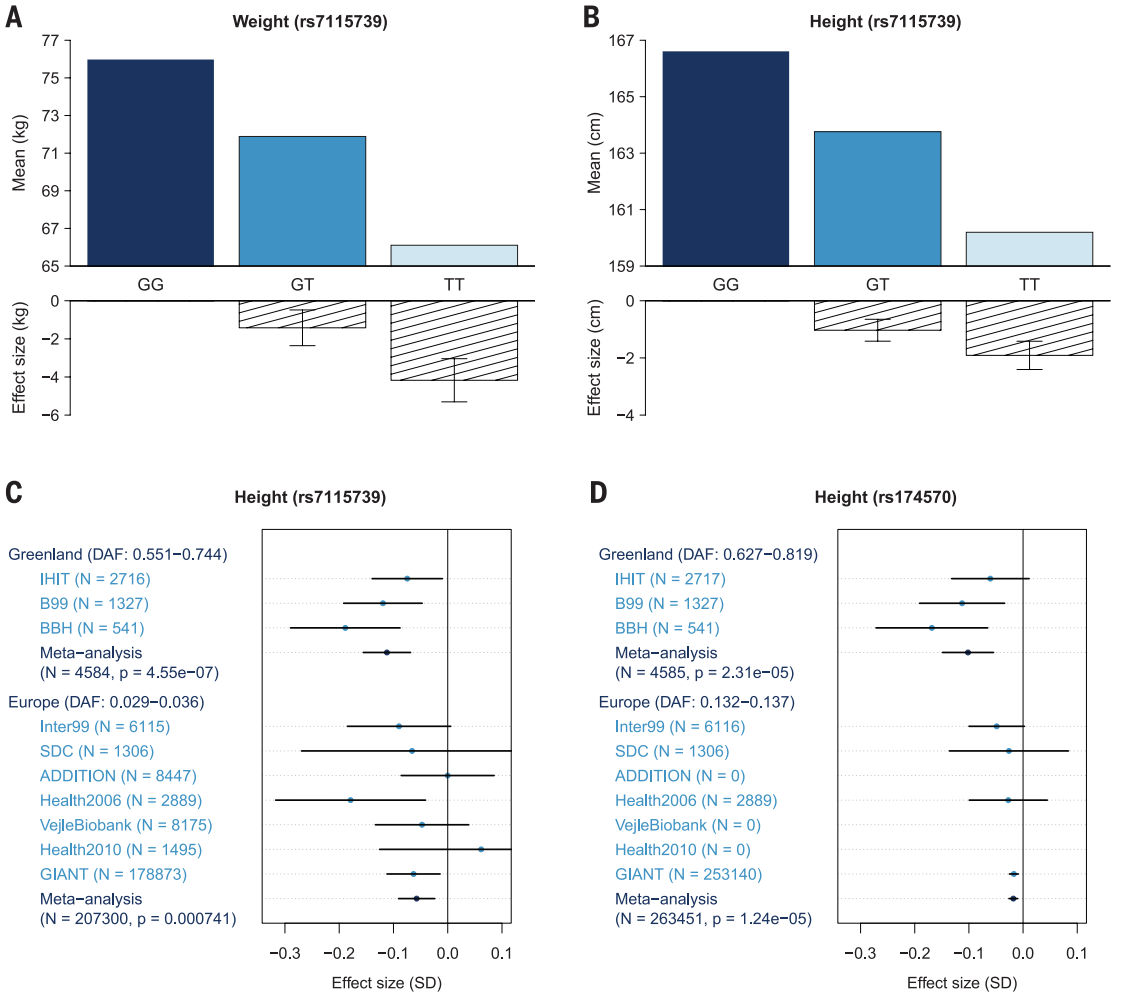


Fig. 2. The effect of rs7115739 and rs174570 on weight and height. (**A**) The effect of rs7115739 on weight in the Greenlandic cohorts IHIT and B99. Shown is the mean value stratified by the genotypes of rs7115739 (top) and the estimated effect of carrying one and two copies of the derived allele, respectively (bottom). The effect-size estimates are adjusted for admixture and other confounding factors and were obtained using a linear mixed model applied to untransformed phenotype measurements. Unlike the estimates in the text and table S10, the estimates shown here were obtained without assuming an additive effect. Error bars, ±1 SE. (**B**) As in (A), but for height. (**C**) Effect sizes for height for the derived allele of rs7115739 in three Greenlandic cohorts and seven European cohorts (SDC, Steno Diabetes Center). Point estimates are shown as points and 95% confidence intervals are shown as horizontal bars. For each of the two geographic regions, the results from a meta-analysis of all the cohorts from the region are also shown. *N* indicates the number of individuals analyzed; DAF, derived allele frequency. For the Greenlandic cohorts, the effect sizes were estimated from height measurements that were quantile-transformed to a standard normal distribution. For the European cohorts, height was analyzed as sex-specific standard (*z*) scores. Hence, the effect sizes from the two geographic regions are not directly comparable. (**D**) As in (C), but for rs174570.

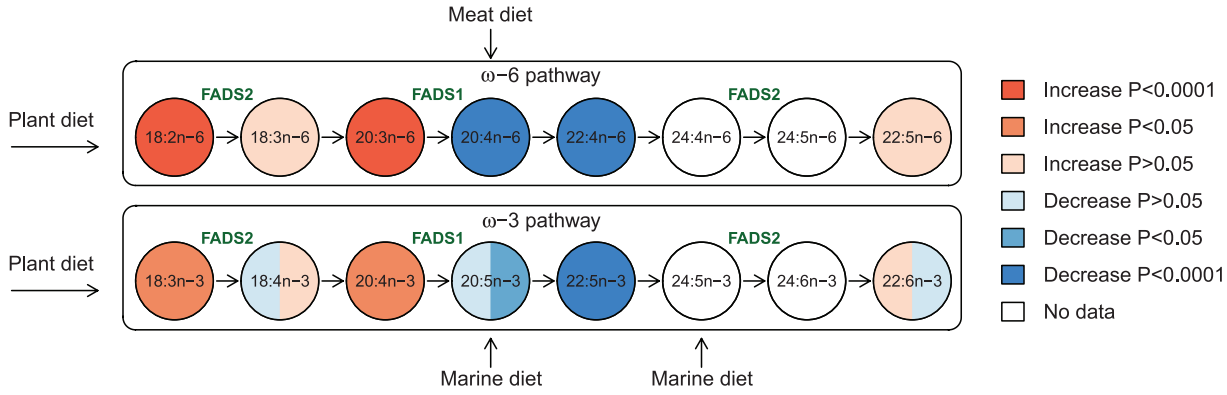


Fig. 3. Results of testing for association between the fatty acids in the omega-3 and omega-6 synthesis pathways and each of two SNPs, rs7115739 and rs174570. The omega-6 (top) and omega-3 (bottom) synthesis pathways are depicted with circles for each fatty acid and arrows for each synthesis step. For each fatty acid, *P* values for tests of association and effect directions for the derived allele are illustrated by the colors on the left (rs7115739) and right (rs174570) halves of the circle. Green text indicates in which of the synthesis steps *FADS1* and *FADS2* play a role. Arrows outside the boxes are simplified indications of where different types of diet enter the two pathways.

because most of the identified SNPs have low allele frequencies in Europeans, they may have been missed by GWAS studies. When combining seven European cohorts, including GIANT (Genetic Investigation of Anthropometric Traits; *26*), we found associations with lower height in carriers of the derived T-allele for rs7115739 (n = 207,300; P = 0.000741) and rs174570 (n = 263,451; $P = 1.24 \times 10^{-5}$) (Fig. 2, C and D, and table S13). The meta-analysis–based effect sizes are equivalent to −0.35 and −0.12 cm for rs7115739 and rs174570, respectively. In contrast, we found no evidence that the six SNPs are associated with weight in Europeans. These results are consistent with results that we obtained when we explicitly tested for differences in effect sizes between Europeans and Greenlandic Inuit (table S14): We found no evidence of a difference in effect size for height for rs7115739 (P = 0.44), but we found significant evidence for a difference in effect size for weight (P = 0.025 and P = 0.012 for rs7115739 and rs174570, respectively), with little or no effect on weight in Europeans. The associations with height in Europeans are unexpected, because this locus was not found to be significant genome-wide in the recent GIANT study of the height of more than 170,000 Europeans (*26*). In addition to the associations with height, we also found known associations with low fasting serum levels of insulin, total cholesterol, and LDL cholesterol for European carriers of low-frequency–derived alleles of *FADS1* variation, suggesting that there may be a protective effect of these variants on cardiometabolic phenotypes (table S13).

To further elucidate the possible functional effects of the alleles of rs7115739 and rs174570, we investigated associations with red blood cell–membrane lipid composition, which reflects fatty-acid intake from the preceding 2 to 4 months and which has previously been measured in IHIT, the largest of our Greenlandic cohorts (*27*). We found significant associations with multiple different fatty acids (fig. S10 and tables S15 and S16). Particularly, we found that the selected alleles are significantly associated with an increase in the concentration of eicosatetraenoic acid (ETA, 20:4n-3) and other omega-3 fatty acids upstream in the omega-3 synthesis pathway, before conversion to EPA (20:5n-3), but a decrease in the concentration of both EPA and omega-3 docosapentaenoic acid (DPA, 22:5n-3), with no significant effect on DHA (22:6n-3) (Fig. 3). These results are consistent with previous observations of linked alleles in Europeans (*28*). The conversion of ETA to EPA is catalyzed by delta-5 desaturases encoded by *FADS1*, and EPA is a major dietary omega-3 fatty acid in the traditional Inuit diet (*18*). Hence, these results suggest that selection affecting the fatty acid desaturases may have compensated for a high dietary intake of EPA.

The changes in the concentration of omega-6 fatty acids mirror those of omega-3 fatty acids (Fig. 3). This might be expected, given that the same enzymes (encoded by *FADS1* and *FADS2*) are involved in both the omega-3 and omega-6 biosynthesis pathways. The similar changes in concentration could therefore be a side effect of selection, driven by a omega-3 PUFA–rich diet. However, selection may also have worked directly on omega-6 fatty acid concentrations early in the ancestral history of Inuit and Native Americans, in the context of a late Paleolithic diet rich in meat from land mammals.

Both rs7115739 and rs174570 show strongly significant associations in conditional analyses where we adjusted for the effects of the other SNP and of rs174602. The remaining three highest-PBS SNPs are in strong LD with rs7115739 in IHIT and would produce similar results. This suggests that there are either multiple causative SNPs or that both rs7115739 and rs174570 are in strong LD with the causal SNP(s).

The challenging environmental conditions of the Arctic have probably imposed strong selective pressures on the Inuit and their ancestors. In all the data that we analyzed, the most pronounced allele-frequency difference between Inuit and other populations was found in a cluster of fatty acid desaturases—*FADS1*, *FADS2*, and *FADS3*—although it is possible that even more extreme differences are present in noncoding regions not covered by our exome data. The *FADS* region has probably been under selection, driven by a diet high in PUFAs. The *FADS* genes have previously been hypothesized to be under selection in other populations in response to dietary changes (*28*, *29*), suggesting that these genes in general play an important role in human adaptation to dietary regimes. Our results also show that genetic variants in fatty acid desaturases have a strong effect on height, probably because of the effect of fatty acid composition and concentration on the regulation of growth hormones (*30*). Previous studies (*31*) have shown that fish oil supplementation is associated with increased concentrations of plasma insulin-like growth factor–1. This study illustrates the utility of evolutionary studies of locally adapted populations for understanding the genetic basis of phenotypic variation among humans.

REFERENCES AND NOTES

1. A. M. Hancock *et al.*, *PLOS Genet.* **7**, e1001375 (2011).
2. F. J. Clemente *et al.*, *Am. J. Hum. Genet.* **95**, 584–589 (2014).
3. I. Moltke *et al.*, *Nature* **512**, 190–193 (2014).
4. H. C. Gulløv, *Grønlands Forhistorie* (Gyldendal, Copenhagen, 2004).
5. Materials and methods are available as supplementary materials on *Science* Online.
6. B. Deutch, J. Dyerberg, H. S. Pedersen, E. Aschlund, J. C. Hansen, *Sci. Total Environ.* **384**, 106–119 (2007).
7. H. O. Bang, J. Dyerberg, H. M. Sinclair, *Am. J. Clin. Nutr.* **33**, 2657–2661 (1980).
8. B. F. Voight *et al.*, *PLOS Genet.* **8**, e1002793 (2012).
9. I. Moltke *et al.*, *Am. J. Hum. Genet.* **96**, 54–69 (2015).
10. D. M. Altshuler *et al.*, *Nature* **467**, 52–58 (2010).
11. X. Yi *et al.*, *Science* **329**, 75–78 (2010).
12. V. S. Voruganti *et al.*, *Front. Genet.* **3**, 86 (2012).
13. L. Xie, S. M. Innis, *J. Nutr.* **138**, 2222–2228 (2008).
14. P. Rzehak *et al.*, *Br. J. Nutr.* **101**, 20–26 (2009).
15. I. M. Heid *et al.*, *Nat. Genet.* **42**, 949–960 (2010).
16. V. Gburcik, W. P. Cawthorn, J. Nedergaard, J. A. Timmons, B. Cannon, *Am. J. Physiol. Endocrinol. Metab.* **303**, E1053–E1060 (2012).
17. F. Collard, G. Delpierre, V. Stroobant, G. Matthijs, E. Van Schaftingen, *Diabetes* **52**, 2888–2895 (2003).
18. A. Jenkinson, M. F. Franklin, K. Wahle, G. G. Duthie, *Eur. J. Clin. Nutr.* **53**, 523–528 (1999).
19. 1000 Genomes Project Consortium, *Nature* **491**, 56–65 (2012).
20. O. Campuzano *et al.*, *Eur. J. Med. Genet.* **56**, 541–545 (2013).
21. K. Mirzaei, A. Hossein-Nezhad, M. Chamari, S. Shahbazi, *Minerva Endocrinol.* **36**, 13–21 (2011).
22. S. Liu *et al.*, *Cell* **157**, 785–794 (2014).
23. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).

24. M. Raghavan *et al*., *Science* **345**, 1255832 (2014).
25. Y. S. Aulchenko *et al*., *Nat. Genet.* **41**, 47–55 (2009).
26. A. R. Wood *et al*., *Nat. Genet.* **46**, 1173–1186 (2014).
27. C. Jeppesen, M. E. Jørgensen, P. Bjerregaard, *Int. J. Circumpolar Health* **71**, 18361 (2012).
28. A. Ameur *et al*., *Am. J. Hum. Genet.* **90**, 809–820 (2012).
29. I. Mathieson *et al*., http://biorxiv.org/content/early/2015/03/13/016477.
30. H.-J. Quabbe, H.-J. Bratzke, U. Siegers, K. Elban, *J. Clin. Invest.* **51**, 2388–2398 (1972).
31. C. T. Damsgaard, C. Mølgaard, J. Matthiessen, S. N. Gyldenløve, L. Lauritzen, *Pediatr. Res.* **71**, 713–719 (2012).

ACKNOWLEDGMENTS

We thank the Greenlandic participants and the funding agencies and research centers that made this study possible: the Human Frontiers in Science Program Organization (grant LT00320/2014); the Danish Council for Independent Research (grant DFF-YDUN); the Villum Foundation; the Steno Diabetes Center; NIH (grant R01-HG003229); the Leverhulme Programme Grant (grant RP2011-R-045); the University of California–Merced startup funds; Karen Elise Jensen's Foundation and NunaFonden, which supported the collection of data from the Greenlandic cohorts; and the Novo Nordisk Foundation Center for Basic Metabolic Research, which is an independent research center at the University of Copenhagen and is partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk). We also thank T. Lauritzen and A. Sandbæk for the use of the ADDITION (Anglo-Danish-Dutch Study of Intensive Treatment In People with Screen Detected Diabetes in Primary Care) cohort. The Vejle Diabetes Biobank was funded by the Danish Medical Research Council and Vejle Hospital. The genotyping and exome sequencing data from this project are available to researchers who have received ethics approval from the Greenland Research Ethics Committee (nun@nanoq.gl) and can be obtained by contacting T.H.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1343/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S17
References (*32*–*67*)

30 March 2015; accepted 17 August 2015
10.1126/science.aab2319

STRUCTURAL BIOLOGY

Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus

Hongrong Liu[1]*† and Lingpeng Cheng[2]*†

Double-stranded RNA (dsRNA) viruses possess a segmented dsRNA genome and a number of RNA-dependent RNA polymerases (RdRps) enclosed in a capsid. Until now, the precise structures of genomes and RdRps within the capsids have been unknown. Here we report the structures of RdRps and associated RNAs within nontranscribing and transcribing cypoviruses (NCPV and TCPV, respectively), using a combination of cryo–electron microscopy (cryo-EM) and a symmetry-mismatch reconstruction method. The RdRps and associated RNAs appear to exhibit a pseudo-D_3 symmetric organization in both NCPV and TCPV. However, the molecular interactions between RdRps and the genomic RNA were found to differ in these states. Our work provides insight into the mechanisms of the replication and transcription in dsRNA viruses and paves a way for structural determination of lower-symmetry complexes enclosed in higher-symmetry structures.

The family Reoviridae causes disease in humans, livestock, insects, and plants. The virions have 10 to 12 segments of dsRNA enclosed in a single-, double-, or triple-layered capsid. The inner capsids (cores) remain intact after the viruses are delivered into the host cell's cytoplasm, and the RNA-dependent RNA polymerases (RdRps) repeatedly transcribe RNA from the minus-strand RNA genome within the core (*1*, *2*). Assembly of the reovirus cores requires encapsidation of the genomic RNA plus strands, along with a roughly equal number of RdRps. The maturation of the reoviruses is accompanied by RdRps-driven synthesis of RNA minus strands complementary to the plus strands, in turn forming genomic double-stranded RNA (dsRNA) segments within the mature virions (*3*, *4*).

Although structures of viral capsids and isolated RdRp complexes have been studied extensively for more than two decades (*3*–*13*), the structures of genomes and RdRps within viral capsids have thus far evaded determination. In this study, we used cryo–electron microscopy (cryo-EM), in combination with our symmetry-mismatch reconstruction method, to report the structures of RdRps and associated RNAs for nontranscribing and transcribing cypoviruses (NCPV and TCPV, respectively) in the family Reoviridae.

Cypovirus particles were isolated and purified, and viral transcription was assayed (*14*, *15*). We reconstructed the structures of the NCPV and TCPV without imposing any symmetry (see supplementary materials and methods). Our analysis of the NCPV showed that the genomic RNAs and RdRps are located inside the capsid within a region of 510 Å radius. The structure of the genomic RNAs is of spherical outline, is composed of regularly distributed layers that are formed by discontinuous dsRNA fragments running in parallel, and is associated with RdRps (Fig. 1, A to C; fig. S1; movie S1). Each RdRp is anchored at the inner surface of the capsid and surrounded by multiple layers of dsRNA (Fig. 1, B and C). The distance between two adjacent dsRNA fragments within the same layer is fixed at ~25 Å, whereas two adjacent layers are ~30 Å apart. The double helices of both dsRNA fragments located close to the inner capsid surface and interacting with the RdRps have a measured helix pitch of ~28 Å (Fig. 1C). The dsRNA fragment structures located closer to the spherical center are not as well resolved as the those at the periphery (fig. S1). Each RdRp density anchors to the inner surface of the capsid, slightly off-center from the fivefold axis (Fig. 1B) (*16*). These RdRps and the associated dsRNA fragments appear to exhibit a pseudo-D_3 symmetric organization (Fig. 1A and figs. S1 to S3), allowing for 12 distinct locations of RdRps inside a viral capsid: Two groups containing three RdRps (threefold RdRps) each approach and are symmetrically arranged about the threefold axes on opposite sides of the virion, and three groups containing two RdRps (twofold RdRps) each approach and are symmetrically arranged about the twofold axes that encircle the center of the virion (fig. S4). Within the three-dimensional density maps, the average density value of the twofold RdRps amounts to approximately two-thirds of the average density value of the threefold RdRps. In contrast, the dsRNA densities surrounding the twofold and the threefold RdRps are all of similar intensity. We reason that this reflects six RdRps occupying the six positions of the threefold RdRps and only four RdRps occupying the six positions of the twofold RdRps (thus, two-thirds of the average density). Therefore, the total number of RdRps within the capsid is 10, in tentative agreement with the observation that each cypovirus genome contains only 10 RNA segments, with each genome segment being specifically associated with one RdRp (*17*). Our structural analysis also revealed that TCPV and NCPV have almost identical genome structures (figs. S2 and S5), except for those genome regions that interact with RdRps. Given the great variations of size and the encoded genes of the 10 different genomic RNA segments in each cypovirus, it is likely that the observed D_3 symmetry in the dsRNA organization does not reflect the true organization of the RNA genome. The layers of the dsRNA fragment resemble the organization of the cholesteric liquid crystal (*18*) (fig. S1 and movie S2), which is consistent with earlier evidence that the dsRNA genome forms liquid crystalline arrays within the highly condensed capsid (*5*). The liquid crystalline model of genome

[1]College of Physics and Information Science, Hunan Normal University, Changsha, Hunan 410081, China. [2]School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work. **†Corresponding author. E-mail: hrliu@hunnu.edu.cn (H.L.); lingpengcheng@mail.tsinghua.edu.cn (L.C.)**

packaging does not support the spool model of Reoviridae genomic organization based on icosahedrally averaged genome structures (*19*–*21*).

Using the presence of D_3 symmetry of the RdRp structures, we further improved the resolution of RdRps in NCPV and TCPV from ~6 Å (Fig. 2, A and B) to ~5 Å (Fig. 2C, fig. S6, table S1, and movie S3) by applying this symmetry to the reconstruction. The well-defined densities of the two RdRp conformers in NCPV and TCPV allowed us to build Cα backbone models (Fig. 2, B and C, and figs. S7 and S8). The structure of the TCPV RdRp is very similar to that of the orthoreovirus RdRp λ3 (*3*). We therefore refer to the domains in the cypovirus RdRp according to the nomenclature of the orthoreovirus λ3. TCPV RdRp has four channels that connect to its central catalytic cavity (Fig. 2D); these channels are used for RNA template entry, nucleoside triphosphate (NTP) entry, template exit, or RNA transcript exit (*3*). An additional density, which we attributed to the homolog of μ2 protein in orthoreovirus, was observed attached to the surface region between the channels for template entry and NTP entry of each RdRp in both NCPV and TCPV (Fig. 2A and figs. S7A and S8A). Consistent with an earlier biochemical analysis suggesting that μ2 might be an NTP instead of an RNA 5′-triphosphatase (RTPase) (*22*), this observation implies that the μ2-homology protein is not an RTPase, as it does not have access to the 5′ end of the transcript.

A dsRNA fragment was observed to bind to each of the 12 RdRp bracelet (C-terminal) domains via nucleotides of the second turn downstream of the dsRNA end in NCPV (Fig. 2, A and B). A caplike structure visible at each of the dsRNA ends appears to be the 5′ end of the plus-strand RNA (i.e., the 3′ end of the minus-strand RNA) (Fig. 2B). In response to binding dsRNA, the bracelet domain of NCPV RdRp shifts compared with that of TCPV: One α helix moves into the transcript exit channel, and another two α helices move into the template exit channel, leading to partial blockage of the two channels (Figs. 2E and 3, A and B).

In TCPV, the end of the dsRNA approaches the template entry channel of RdRp (Fig. 4, A and B). The dsRNA is unwound at the channel entrance, with one strand (presumably the transcriptional template; i.e., the 3′ end of the minus-strand RNA) inserting into the channel and reaching the RdRp active site, and the other strand (the capped 5′ end of plus-strand RNA) being tethered to the RdRp surface (Fig. 4, B and C). Fitting the orthoreovirus RdRp atomic model into the TCPV RdRp density reveals that TCPV RdRp's plus-strand RNA binding site superimposes with the proposed RdRp cap recognition site (*3*) (Fig. 4D). No RNA density was observed inside the template exit channel.

The initiation stage of the transcription is known to be less efficient than the elongation process (*23*). Thus, a considerable percentage of the RdRps in our sample are initiating transcription, which suggests that the TCPV RdRp structure might represent the initiation stage of transcription. We did not observe any auxiliary proteins at the entrance of the template entry channel, suggesting that this RdRp, like the polymerases from bacteriophage Φ6 and bursal disease virus (*13*, *24*), is able to separate the two RNA strands in the absence of additional helicases.

We also observed a number of density features in the active site of TCPV and NCPV RdRps. Some density features, which can be assigned to the template RNA and an NTP, were specifically observed in the active site of TCPV RdRp. Models of the first four nucleotides of the template RNA (3′-UCAU) and an NTP can be built into the densities (Fig. 4E and fig. S9). The channels for transcript and template exit, which are partially blocked in the NCPV RdRp, become open in the TCPV RdRp (Fig. 3). These conformational changes are likely required for the NCPV RdRp's switch to transcription mode. In addition, we observed that a loop present in a retracted conformation in the NCPV RdRp becomes extended in TCPV (Fig. 4F), narrowing the template exit channel; this observation suggests that the transcript will collide with this loop during transcription (Figs. 2D and 4E). Thus, the loop might prevent the transcript from entering the template exit channel and may instead direct it toward the transcript exit channel during transcription. Presumably, this loop (the "switch loop") is also involved in replication, where it is in the retracted conformation, as shown for the NCPV RdRp, thus allowing nascent dsRNA to egress from the template exit channel.

On the basis of previous structural studies (*3*, *4*, *15*, *16*, *25*) and our RdRp and RNA structures, we propose a model for dsRNA virus replication and transcription. The replication process

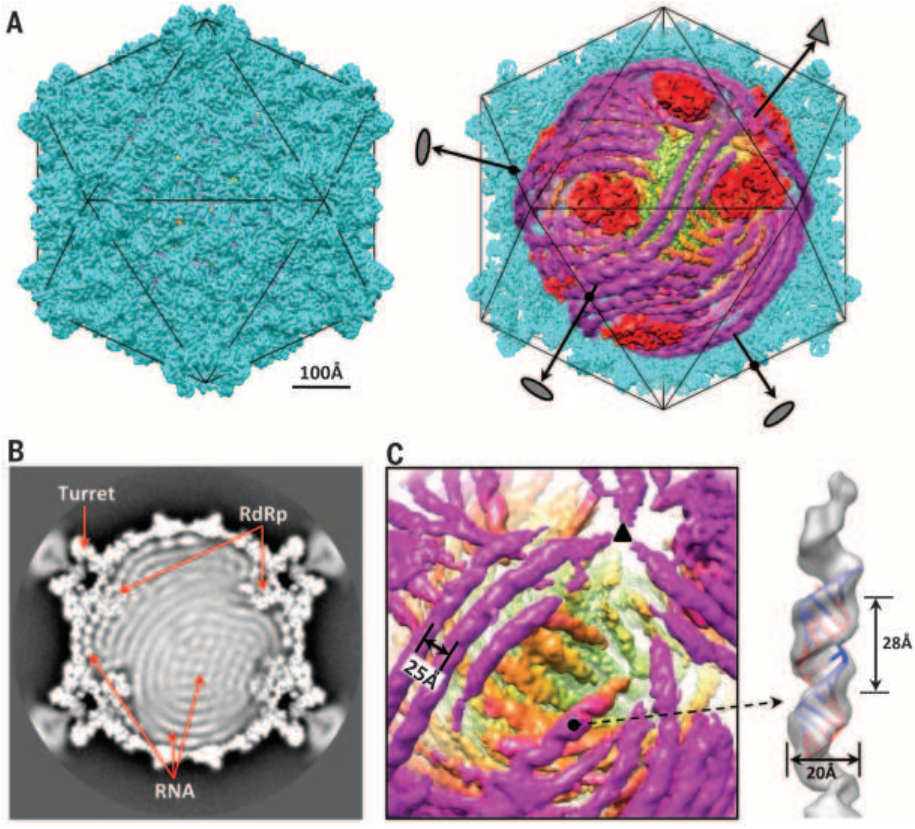


Fig. 1. Structures of the RNA fragments and RdRps within the NCPV capsid. (**A**) (Left) NCPV icosahedral capsid structure. (Right) Half of the icosahedral capsid (blue) is removed to show the structures of genomic dsRNA and RdRps (red). One threefold axis and three twofold axes of pseudo-D_3 symmetry are indicated by a triangle and ellipses, respectively. (**B**) Central slice of the NCPV structure showing the genomic dsRNA and RdRps. (**C**) (Left) Zoomed-in view of the dsRNA. The RdRp has been removed for clarity. (Right) Zoomed-in view of a dsRNA fragment.

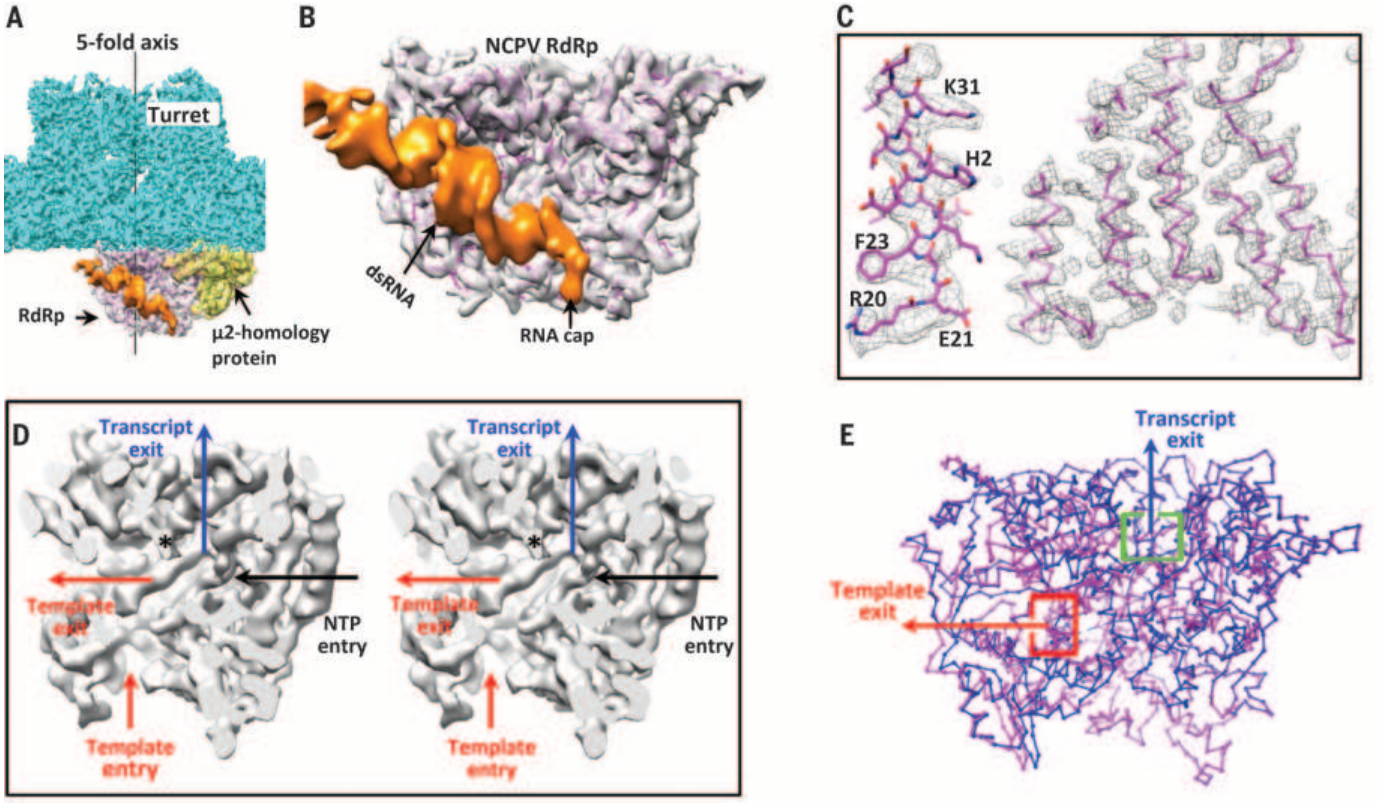


Fig. 2. Structures of the RdRp complexes within NCPV and TCPV. (**A**) Location of an RdRp within the NCPV capsid (partially shown). The RdRp density map (gray) is superimposed on its Cα model (magenta). (**B**) Zoomed-in view of the NCPV RdRp in (A). The μ2-homology protein is removed for clarity. (**C**) Partial view of the density map (mesh) of the RdRp superimposed on its Cα model. Several N-terminal α helix residues (left) are shown (National Center for Biotechnology Information accession number AF323782). K, Lys; H, His; F, Phe; R, Arg; E, Glu. (**D**) Cut-open view (cross-eye stereo view) showing the four channels in the TCPV RdRp. Asterisks denote the location of the switch loop. (**E**) Cα model of the NCPV RdRp (magenta) superimposed on that of the TCPV RdRp (blue).

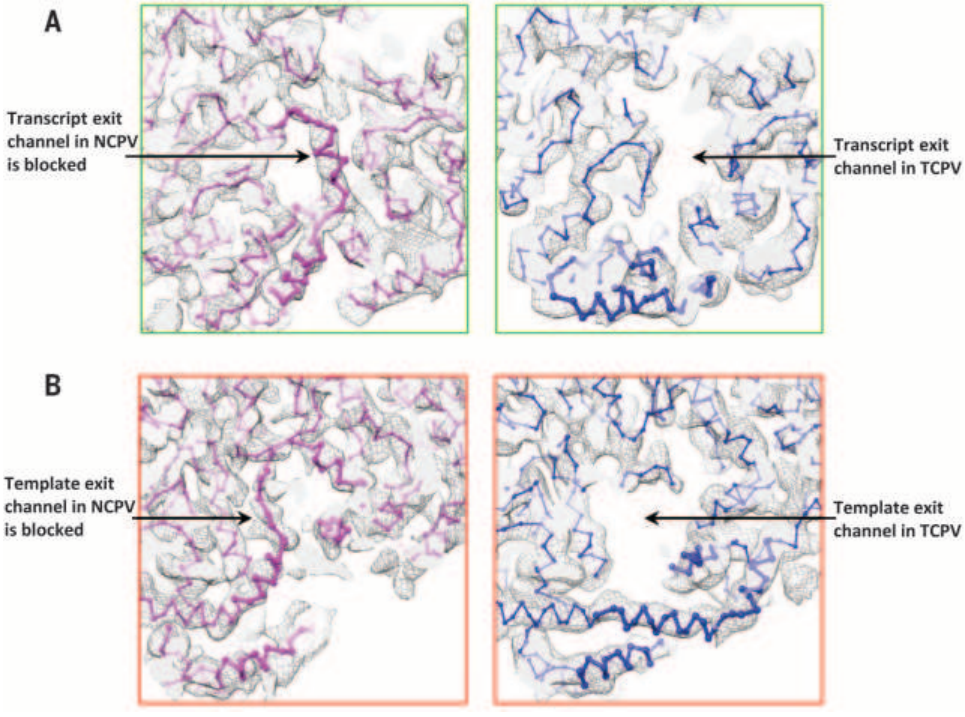


Fig. 3. Conformational changes between RdRps of NCPV and TCPV. Cα models are shown in bold. (**A**) Zoomed-in view of the transcript exit channels in the NCPV and TCPV RdRps shown in Fig. 2E. The transcript exit channel in NCPV is blocked by an α helix from the bracelet domain. (**B**) Zoomed-in view of the template exit channels in NCPV and TCPV RdRps. The template exit channel in NCPV is blocked by two α helices from the bracelet domain.

is initiated from the 3′ end of the plus-strand RNA (template), and the plus-strand RNA is pulled through the RdRp via the template entry channel (fig. S10A). The dsRNA product exits through the template exit channel (*3*) and finally binds to the bracelet domain of the RdRp (Fig. 2B and fig. S10B). In NCPV, the interaction between the dsRNA and RdRp keeps the 3′ end of minus-strand RNA in close proximity to the template entry channel so that this end will be ready to act as the transcriptional template (Fig. 2B and fig. S10). Upon transcription initiation, the end of the dsRNA is released from the binding site, which triggers the opening of the transcript and template exit channels, as well as the extension of the switch loop. The released dsRNA is unwound at the entrance of the template entry channel (Fig. 4B), a process that is likely to be catalyzed by the RdRp itself. The 3′ end of the minus strand inserts into the channel and serves as the RNA template, whereas the cap of the 5′ end of the plus strand binds to the RdRp (*3*). The nascent RNA duplex is split, and the transcript is then directed toward the transcript exit channel by an unknown mechanism, possibly controlled by the switch loop (*3*) (fig. S10C). The template RNA re-anneals with the plus-strand RNA tethered to the RdRp after it emerges from the template exit channel (*3*) (fig. S10D). This will ensure that the 3′ end of the template stays at the entrance of the template entry channel, facilitating efficient insertion of the 3′ end into the template channel during the next round of transcription (*3*). The transcript exit channel opens toward a peripentonal channel in the capsid shell, from which the transcript exits the RdRp before interaction with 5′ capping enzymes (*3*, *4*, *15*, *16*, *25*).

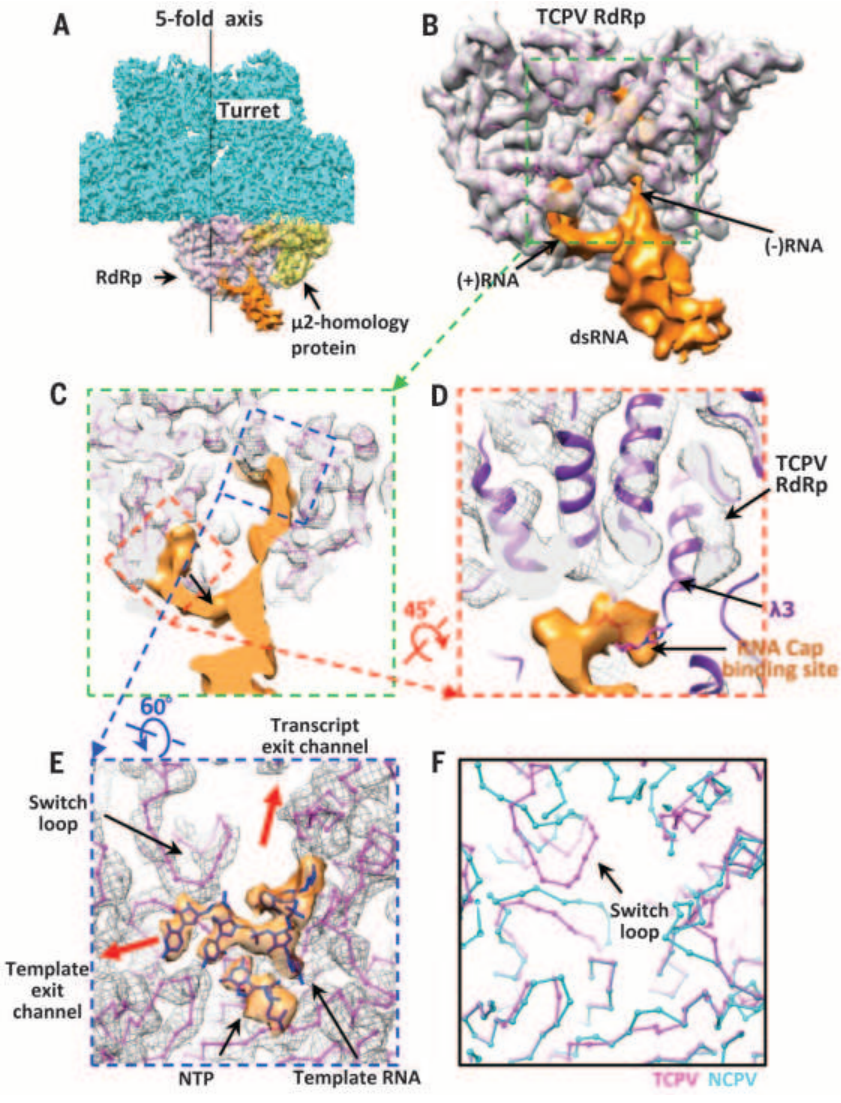


Fig. 4. TCPV RdRp (gray) interacting with RNA (orange). (**A**) Location of an RdRp within the TCPV capsid (partially shown). The RdRp density map (gray) is superimposed on its Cα model (magenta). (**B**) Zoomed-in view of the TCPV RdRp in (A). The dsRNA is shown unwound at the entrance of the RdRp template entry channel. (**C**) Cut-open view showing the inside of the green box indicated in (B). (**D**) Zoomed-in view of the cap binding site of the RdRp. The atomic model of orthoreovirus λ3 (purple) is superimposed on the density map of RdRp, showing that the cap recognition site in the orthoreovirus λ3 coincides with the RNA binding site of the RdRp. (**E**) Cut-open view of template RNA and an NTP. (**F**) The switch loop, which is presented in a retracted conformation in NCPV RdRp (blue), becomes extended in TCPV RdRp (magenta).

REFERENCES AND NOTES

1. J. A. Lawton, M. K. Estes, B. V. V. Prasad, *Nat. Struct. Biol.* **4**, 118–121 (1997).
2. P. Mertens, *Virus Res.* **101**, 3–13 (2004).
3. Y. Tao, D. L. Farsetta, M. L. Nibert, S. C. Harrison, *Cell* **111**, 733–745 (2002).
4. X. Lu *et al.*, *Structure* **16**, 1678–1688 (2008).
5. K. A. Dryden *et al.*, *J. Cell Biol.* **122**, 1023–1041 (1993).
6. B. Böttcher, S. A. Wynne, R. A. Crowther, *Nature* **386**, 88–91 (1997).
7. J. F. Conway *et al.*, *Nature* **386**, 91–94 (1997).
8. K. M. Reinisch, M. L. Nibert, S. C. Harrison, *Nature* **404**, 960–967 (2000).
9. N. G. Abrescia *et al.*, *Nature* **432**, 68–74 (2004).
10. S. Duquerroy *et al.*, *EMBO J.* **28**, 1655–1665 (2009).
11. L. Li, J. Jose, Y. Xiang, R. J. Kuhn, M. G. Rossmann, *Nature* **468**, 705–708 (2010).
12. H. Liu *et al.*, *Science* **329**, 1038–1043 (2010).
13. S. J. Butcher, J. M. Grimes, E. V. Makeyev, D. H. Bamford, D. I. Stuart, *Nature* **410**, 235–240 (2001).
14. L. Cheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1373–1378 (2011).
15. C. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6118–6123 (2012).
16. X. Zhang, S. B. Walker, P. R. Chipman, M. L. Nibert, T. S. Baker, *Nat. Struct. Biol.* **10**, 1011–1018 (2003).
17. K. Yazaki, K. Miura, *Virology* **105**, 467–479 (1980).
18. P. G. de Gennes, J. Prost, *The Physics of Liquid Crystals* (Oxford Univ. Press, Oxford, 1993).
19. P. Gouet *et al.*, *Cell* **97**, 481–490 (1999).
20. J. B. Pesavento, J. A. Lawton, M. K. Estes, B. V. V. Prasad, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1381–1386 (2001).
21. Q. Xia, J. Jakana, J. Q. Zhang, Z. H. Zhou, *J. Biol. Chem.* **278**, 1094–1100 (2003).
22. J. Kim, J. S. Parker, K. E. Murray, M. L. Nibert, *J. Biol. Chem.* **279**, 4394–4403 (2004).
23. E. V. Makeyev, J. M. Grimes, *Virus Res.* **101**, 45–55 (2004).
24. J. Pan, V. N. Vakharia, Y. J. Tao, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7385–7390 (2007).
25. B. Zhu *et al.*, *J. Mol. Biol.* **426**, 2167–2174 (2014).

ACKNOWLEDGMENTS

We thank H. Wang for critical discussion of the manuscript, B. Zhu and X. Li for assistance with computer programming, X. Gong for help with model refinement, T. Juelich for language editing, C. Yang for discussion on sample preparation, G. Ji and X. Huang for assistance with cryo-EM, and P. Zhu for support. This research was supported by the National Natural Science Foundation of China (91230116, 31170697, 31370736, and 31000333), the National Basic Research Program of China (2010CB912403), the Program for New Century Excellent Talents (NCET-13-0787), and the Hunan Provincial NSF (13JJ1017). All EM data were collected at the Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences. We declare no competing financial interests. The electron density maps and Cα models have been deposited in the EM Data Bank and the Protein Data Bank (PDB) under EM accession numbers EMD-6321 and EMD-6322 and PDB IDs 3JA4 and 3JA5, respectively. L.C. and H.L. designed the reconstruction algorithm, analyzed the data, and contributed to the writing of the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1347/suppl/DC1
Materials and Methods
Figs. S1 to S15
Table S1
References (*26*–*41*)
Movies S1 to S3

15 December 2014; accepted 12 August 2015
10.1126/science.aaa4938

PROSTATE CANCER

RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance

David T. Miyamoto,[1,2*] Yu Zheng,[1,3*] Ben S. Wittner,[1,4] Richard J. Lee,[1,4] Huili Zhu,[1] Katherine T. Broderick,[1] Rushil Desai,[1] Douglas B. Fox,[1] Brian W. Brannigan,[1] Julie Trautwein,[1] Kshitij S. Arora,[1,5] Niyati Desai,[1,5] Douglas M. Dahl,[1,6] Lecia V. Sequist,[1,4] Matthew R. Smith,[1,4] Ravi Kapur,[7] Chin-Lee Wu,[1,5] Toshi Shioda,[1] Sridhar Ramaswamy,[1,4] David T. Ting,[1,4] Mehmet Toner,[7] Shyamala Maheswaran,[1,8]† Daniel A. Haber[1,3,4]†

Prostate cancer is initially responsive to androgen deprivation, but the effectiveness of androgen receptor (AR) inhibitors in recurrent disease is variable. Biopsy of bone metastases is challenging; hence, sampling circulating tumor cells (CTCs) may reveal drug-resistance mechanisms. We established single-cell RNA-sequencing (RNA-Seq) profiles of 77 intact CTCs isolated from 13 patients (mean six CTCs per patient), by using microfluidic enrichment. Single CTCs from each individual display considerable heterogeneity, including expression of *AR* gene mutations and splicing variants. Retrospective analysis of CTCs from patients progressing under treatment with an AR inhibitor, compared with untreated cases, indicates activation of noncanonical Wnt signaling ($P = 0.0064$). Ectopic expression of Wnt5a in prostate cancer cells attenuates the antiproliferative effect of AR inhibition, whereas its suppression in drug-resistant cells restores partial sensitivity, a correlation also evident in an established mouse model. Thus, single-cell analysis of prostate CTCs reveals heterogeneity in signaling pathways that could contribute to treatment failure.

After the initial response of metastatic prostate cancer to androgen deprivation therapy (ADT), it invariably recurs as castration-resistant disease (*1*). Second-line inhibitors of the androgen receptor (AR) have been shown to increase overall survival in castration-resistant prostate cancer (CRPC), consistent with the reactivation of AR signaling in the tumor, but responses are heterogeneous and often short-lived, and resistance to therapy is a pressing clinical problem (*1*). In other types of cancer, molecular analyses of serial biopsies have enabled the study of acquired drug-resistance mechanisms, intratumor heterogeneity, and tumor evolution in response to therapy (*2*)—an approach that is restricted by the predominance of bone metastases in prostate cancer (*3*, *4*). Thus, isolation of circulating tumor cells (CTCs) may enable noninvasive monitoring, as patients initially respond and subsequently become refractory to therapies targeting the AR pathway (*5*). Here, we established single-cell RNA-sequencing (RNA-Seq) profiles of CTCs, individually isolated after microfluidic enrichment from blood specimens of men with prostate cancer, to address their heterogeneity within and across different patients and their differences from primary tumor specimens. Retrospective analyses of clinical and molecular data were then performed to identify potentially clinically relevant mechanisms of acquired drug resistance.

Building on earlier approaches for capturing and scoring CTCs (*3*), highly efficient microfluidic technologies enable molecular analyses (*6*–*9*). We applied the CTC-iChip to magnetically deplete normal hematopoietic cells from whole-blood specimens (*10*). Untagged and unfixed CTCs were identified by cell surface staining for epithelial and mesenchymal markers [epithelial cell adhesion molecule (EpCAM) and cadherin-11 (CDH11), respectively], and absent staining for the common leukocyte marker CD45. These labeled CTCs were then individually micromanipulated (fig. S1, A and B). A total of 221 single-candidate prostate CTCs were isolated from 18 patients with metastatic prostate cancer and 4 patients with localized prostate cancer (fig. S1C and table S1). Of these, 133 cells (60%) had RNA of sufficient quality for amplification and next-generation RNA sequencing, and 122 (55%) had >100,000 uniquely aligned sequencing reads (*11*) (figs. S1C and S2A). Although many cancer cells in the circulation appear to undergo apoptosis, the presence of intact RNA identifies the subset enriched for viable cells. In addition to candidate CTCs, we also obtained comprehensive transcriptomes for bulk primary prostate cancers from a separate cohort of 12 patients (macrodissected for >70% tumor content) (table S2), 30 single cells derived from four different prostate cancer cell lines, and five patient-derived leukocyte controls (fig. S1C). The leukocytes were readily distinguished by their expression of hematopoietic lineage markers and served to exclude any CTCs with potentially contaminating signals. Strict expression thresholds were used to define lineage-confirmed CTCs, scored by prostate lineage-specific genes (*PSA*, *PSMA*, *AMACR*, and *AR*) and standard epithelial markers (*KRT7*, *KRT8*, *KRT18*, *KRT19*, and *EPCAM*) (*11*) (fig. S2B). Given the presence of leukocyte transcripts suggestive of cellular contamination or misidentification during selection, 28 cells were excluded, and, given low expression of both prostate lineage-specific genes and standard epithelial markers, 17 cells were excluded. The remaining 77 cells (from 13 patients; average of six CTCs per patient) were defined as categorical CTCs (fig. S1C and table S1).

Unsupervised hierarchical clustering analysis of single prostate CTCs, primary tumor samples, and cancer cell lines resulted in their organization into distinct clusters (Fig. 1A). Single CTCs from an individual patient showed considerably greater intercellular heterogeneity in their transcriptional profiles than single cells from prostate cancer cell lines (Fig. 1, B and C) (mean correlation coefficient 0.10 versus 0.44, $P < 1 \times 10^{-20}$), but they strongly clustered according to patient of origin, which indicated higher diversity in CTCs from different patients (Fig. 1C and fig. S2C) (mean correlation coefficient 0.10 for CTCs within patient versus 0.0014 for CTCs between patients, $P = 2.0 \times 10^{-11}$).

We examined gene markers of prostate lineage, epithelial, mesenchymal, and stem cell fates, and cellular proliferation (Fig. 2A). Epithelial markers were abundantly expressed [>10 reads per million (rpm)] by nearly all CTCs analyzed (92%), whereas mesenchymal genes were not up-regulated compared with primary tumors or prostate cancer–derived cell lines. Among robustly expressed transcripts were putative stem cell markers (*12*), including *ALDH7A1*, *CD44*, and *KLF4*, present in 60% of CTCs. In addition, 47% of CTCs expressed markers of cell proliferation. We performed differential gene expression analysis to identify genes that are up-regulated in prostate CTCs compared with primary tumor samples. A total of 711 genes were highly expressed in CTCs compared with primary tumors; the most enriched were (i) the molecular chaperone *HSP90AA1*, which regulates the activation and stability of AR, among other functions (*13*), and (ii) the non-coding RNA transcript *MALAT1*, which has been implicated in alternative mRNA splicing and transcriptional control of gene expression (*14*) (Fig. 2B, fig. S4A, and table S3) [false discovery rate (FDR) < 0.1, and fold change > 2]. We used the Pathway Interaction Database (PID) (*15*) to identify key molecular pathways up-regulated in

[1]Massachusetts General Cancer Center, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [2]Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [3]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA. [4]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [5]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [6]Department of Urology, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [7]Center for Bioengineering in Medicine, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA. [8]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: haber@helix.mgh.harvard.edu (D.H.); smaheswaran@mgh.harvard.edu (S.M.)**

CTCs versus primary tumors, as well as those upregulated in metastatic versus primary prostate tumors, on the basis of analyses of previously published data sets (*11*) (Fig. 2C, fig. S5, and table S4). In total, 21 pathways were specifically enriched in prostate CTCs, with the majority implicated in growth factor, cell adhesion, and hormone signaling (Fig. 2D and fig. S5).

The AR pathway constitutes the primary therapeutic target in prostate cancer, with specific mutations in *AR* (*1*, *16*) and *AR* mRNA splice variants (*17*, *18*) implicated in acquired resistance. The *AR* transcript was expressed (>10 rpm) in 60 out of 77 (78%) CTCs (12 out of 13 patients with prostate cancer). The T877A mutation (Thr^{877} replaced by Ala) in AR, previously associated with ligand promiscuity and resistance to antiandrogens (*1*), was identified in five out of nine CTCs from a single (1 out of 13) patient with metastatic CRPC (Fig. 3A and table S5). The F876L mutation (Phe^{876} replaced by Leu) in the ligand-binding domain, which converts the AR antagonist enzalutamide to a potential AR agonist (*19*, *20*), was not detected in any of the CTCs (<1 out of 32 CTCs with sufficient sequencing reads for mutational analysis). Thus, in our study, point mutations in *AR* known to be associated with altered signaling were uncommon in patients with CRPC, consistent with other reports (*4*, *21*).

We then analyzed *AR* mRNA splice variants lacking a ligand-binding domain and encoding constitutively active proteins (*1*, *17*). These alternative transcripts are not attributable to discrete genetic mutations, but they are commonly expressed in CRPC (*4*), and detection in bulk CTC preparations of the single splice variant *AR-V7* has been correlated to clinical resistance to antiandrogens (*18*). Our single-cell analysis revealed far more complex and heterogeneous patterns of *AR* splice-variant expression among individual CTCs from patients with CRPC: 33 out of 73 (43%) expressed at least one type of *AR* splice variant (8 out of 11 CRPC patients). Among these CTCs, 26 out of 73 (36%) expressed *AR-V7* (8 out of 11 patients); 18 out of 73 (25%) had a distinct splice form *ARv567es* (*AR-V12*) (8 out of 11 patients); and 7 out of 73 (10%) had *AR-V1*, *AR-V3*, or *AR-V4* splice variants (5 out of 11 patients), all of which are known to result in altered signaling (Fig. 3A and table S6). Simultaneous expression of more than one type of *AR* splice variant was observed in 13 out of 73 (18%) single CTCs (7 out of 11 patients). In total, 7 out of 11 (64%) CRPC patients had CTCs with more than one type of *AR* alteration (including *AR* splice variants and point mutations). In contrast, no such alterations were evident in 12 primary prostate tumors, and only one out of four CTCs from two patients with castration-sensitive prostate cancer (CSPC) that was previously untreated had low-level expression of the *AR-V7* splice variant (Fig. 3A and table S6). Aberrant alternative splicing is a recognized feature of many cancers (*22*), and indeed, another prostate-specific transcript, *KLK3* (*PSA*) (*23*), showed many more alternative splice variants in CTCs from metastatic patients compared with primary tumors (P = 0.0088) (fig. S4B). Taken together, our observations indicate that intrapatient tumor heterogeneity is such that individual CTCs may have different or multiple mRNA splicing alterations.

Tumor heterogeneity is thought to increase further as second-line therapies exert additional selective pressure. We performed retrospective differential analyses in subsets of CTCs to identify mechanisms of resistance to enzalutamide, a potent AR inhibitor recently approved by the U.S. Food and Drug Administration for CRPC (*24*). From eight patients with metastatic prostate cancer who had not received enzalutamide (group A), 41 CTCs were compared with 36 CTCs from five patients whose cancer exhibited radiographic and/or prostate-specific antigen (PSA) progression during therapy (group B) (Fig. 3A and table S1). Gene set enrichment analysis (GSEA) of candidate PID cellular signaling pathways showed significant enrichment for noncanonical Wnt signaling in group B compared with group A CTCs

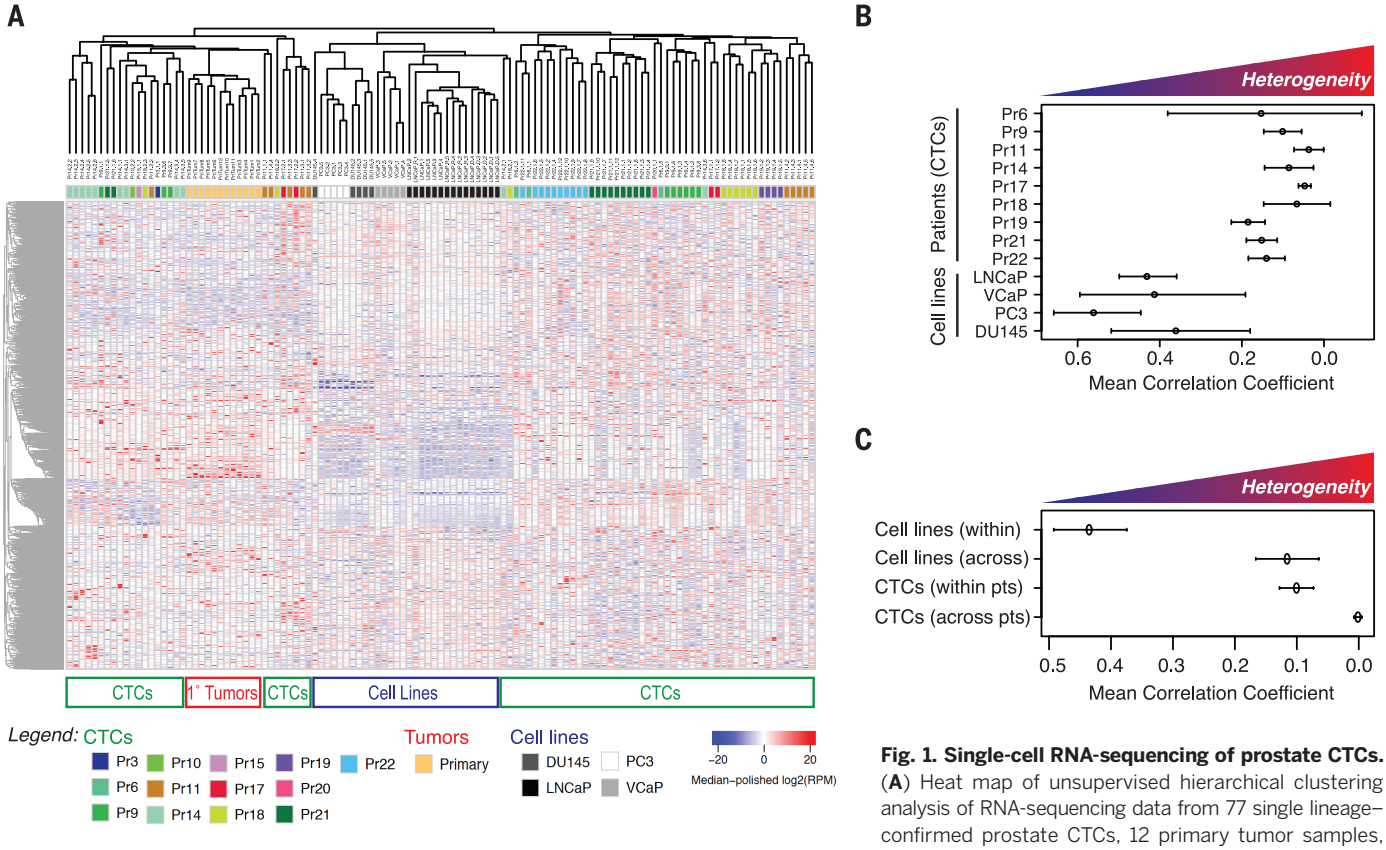
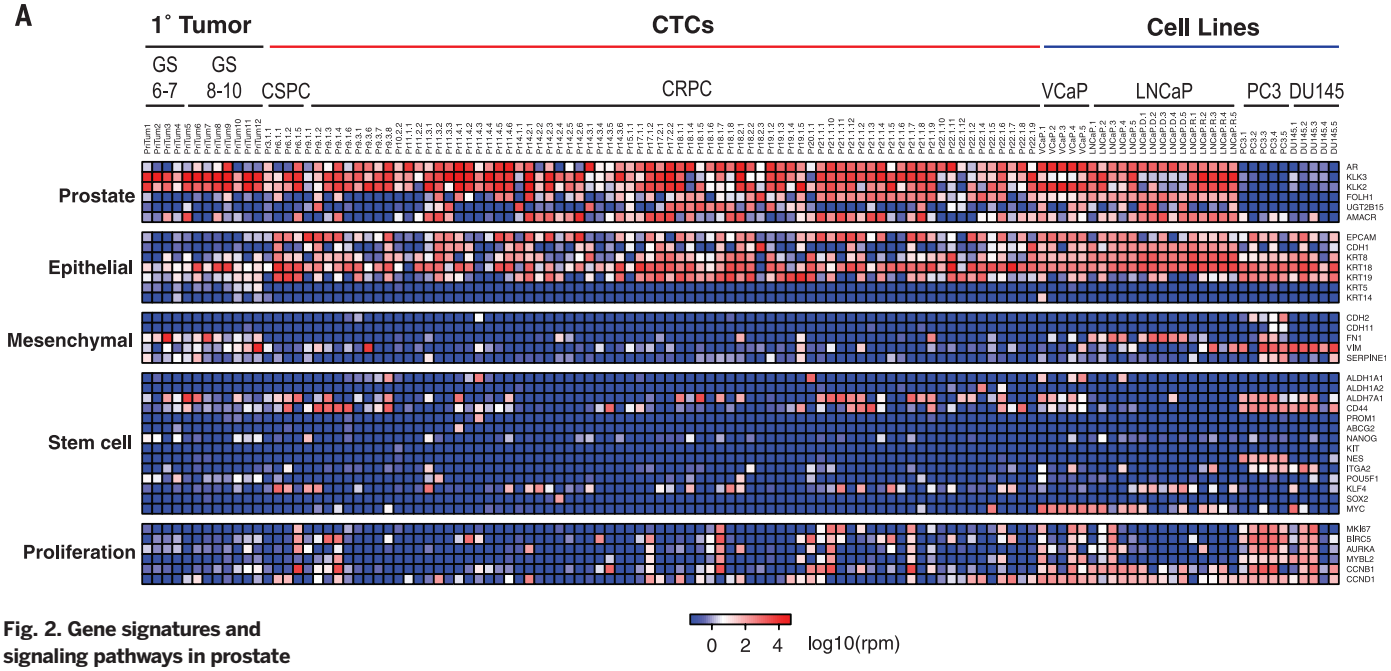
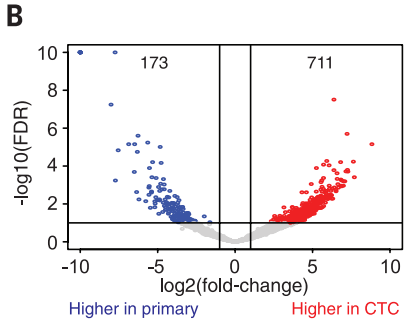


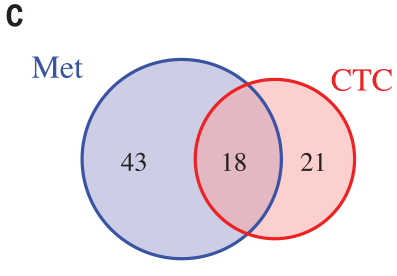
Fig. 1. Single-cell RNA-sequencing of prostate CTCs. (**A**) Heat map of unsupervised hierarchical clustering analysis of RNA-sequencing data from 77 single lineage–confirmed prostate CTCs, 12 primary tumor samples, and 30 single cells from four prostate cancer cell lines. (**B**) Heterogeneity, measured by mean correlation coefficient within individual samples with three or more cells available for analysis. (**C**) Heterogeneity analysis showing mean correlation coefficients from expression data for CTCs between and within patients (0.0013838 versus 0.10055; Holm corrected $P = 2.0 \times 10^{-11}$), and for prostate cancer cell lines between and within lines (0.11568 versus 0.43534; Holm corrected $P = 5.42 \times 10^{-14}$).

(Fig. 3B and fig. S6A) (P = 0.0064; FDR = 0.239). This signaling pathway, activated by a subset of Wnt ligands, mediates multiple downstream regulators of cell survival, proliferation, and motility (fig. S6B) (*25*–*28*). A separate analysis using a metagene for the PID noncanonical Wnt signature (*11*) (table S7) confirmed enrichment of the signature in group B compared with group A CTCs, at the level of both individual CTCs and individual patients (Fig. 3A) [P = 0.0041 (CTCs); P = 0.04 (patients)]. Among the downstream components of noncanonical Wnt, the most significantly enriched were *RAC1*, *RHOA*, and *CDC42*, signaling molecules involved in actin cytoskeleton remodeling and cell migration (Fig. 3A and fig. S6B) [$P = 1 \times 10^{-6}$ (*RAC1*), P = 0.0046 (*RHOA*), P = 0.0097 (*CDC42*)]. In contrast, *AR* abnormalities were not significantly increased among either individual CTCs or patients, when comparing enzalutamide-resistant versus enzalutamide-naïve cases, using a similar analysis (Fig. 3A).

Although most studies of CRPC have focused on acquired *AR* gene abnormalities, an alternative pathway, glucocorticoid receptor (GR) signaling, has recently been shown to contribute to antiandrogen resistance in a prostate cancer mouse xenograft model (*29*). Within our human prostate CTC data set, *GR* transcripts and a metagene signature of GR signaling (*11*) (table S7) did not reach statistical significance between patients in group A versus B [P = 0.35 (CTCs); P = 0.59 (patients)] (Fig. 3A), but an inverse relationship between *GR* expression and noncanonical Wnt signaling was evident. Among CTCs with low *GR* expression, GSEA analysis showed significant enrichment for noncanonical Wnt signaling in







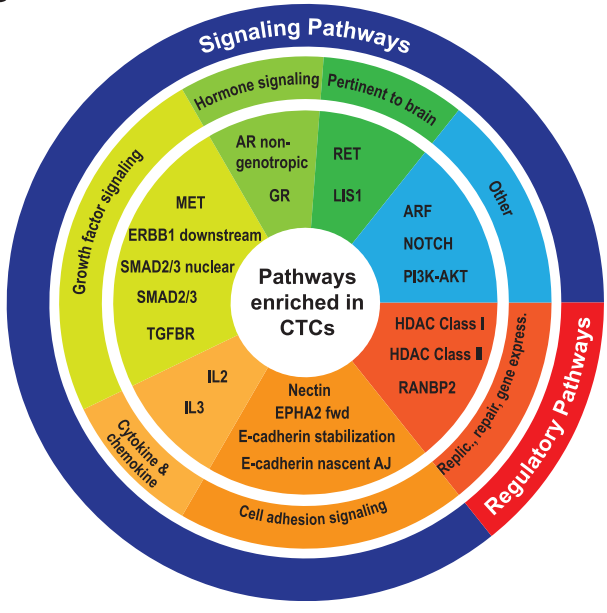
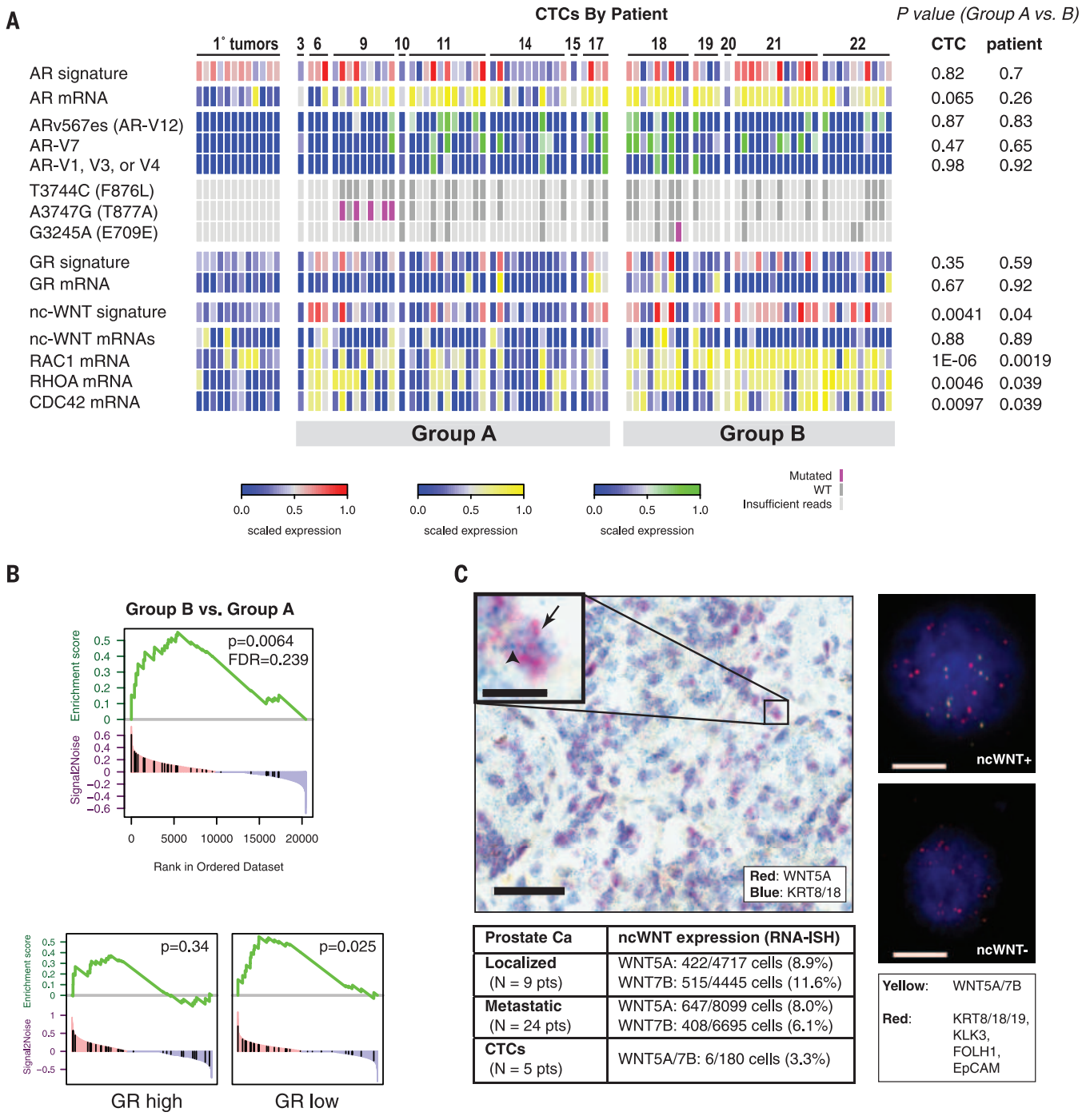


Fig. 2. Gene signatures and signaling pathways in prostate CTCs. (**A**) High-resolution heat map showing expression of selected gene panels in single prostate CTCs, primary tumor samples, and prostate cancer cell lines. GS, Gleason score; VCaP, a PSA-producing cell line; LNCaP.R, LNCaP cells treated with R1881; LNCaP.D, LNCaP cells treated with dimethyl sulfoxide as a vehicle control. (**B**) Genes differentially expressed by prostate CTCs and primary prostate tumors (FDR < 0.1, and fold change > 2). (**C**) PID molecular pathways (*15*) enriched in CTCs compared with primary tumors and in metastases compared with primary tumors (based on analysis of multiple data sets; see fig. S5 and table S4). (**D**) Signaling pathways enriched in prostate CTCs. Molecular pathways from the PID up-regulated in CTCs versus primary tumors (excluding those enriched in metastases compared with primary tumors), organized by PID categorization (*15*) (fig. S5). Abbreviations (other than proteins, clockwise from top): HDAC, histone deacetylase; AJ, adherens junction; IL2 and IL3, interleukins; ERBB1, epidermal growth factor receptor B1; TGFBR, transforming growth factor-β receptor.



Prostate Ca	ncWNT expression (RNA-ISH)
Localized (N = 9 pts)	WNT5A: 422/4717 cells (8.9%) WNT7B: 515/4445 cells (11.6%)
Metastatic (N = 24 pts)	WNT5A: 647/8099 cells (8.0%) WNT7B: 408/6695 cells (6.1%)
CTCs (N = 5 pts)	WNT5A/7B: 6/180 cells (3.3%)

Fig. 3. Heterogeneity of treatment resistance mechanisms in prostate CTCs. (**A**) Heat map depicting androgen receptor (AR) abnormalities, selected signaling pathway signatures, and genes in radical prostatectomy specimens, prostate CTCs from enzalutamide-naïve patients (group A), and prostate CTCs from patients who had radiographic or biochemical progression of disease while receiving treatment with enzalutamide (group B). Noncanonical Wnt signature is from reference (*15*), glucocorticoid receptor (GR) signature is from reference (*29*), and AR signature is from reference (*31*) (table S7). Numbers at top of heat map represent ID numbers (Pr numbers) for patients from which each CTC is derived. (**B**) (Top) GSEA plots showing enrichment of noncanonical Wnt (nc-Wnt) pathway in CTCs from group B (patients with cancer progression on enzalutamide) compared with group A (enzalutamide-naïve patients). (Bottom) Enrichment of noncanonical Wnt pathway in CTCs from group B compared with group A, stratified by GR gene expression. (**C**) (Left) Representative micrograph (40×) of RNA-in situ hybridization assay in metastatic prostate tumors, probing for *WNT5A* and *KRT8/18*, scale bar, 50 μm. (Inset) High magnification, arrow points to *WNT5A* signal (red dot), arrowhead points to *KRT8/18* signal (blue dot), scale bar, 10 μm. Adjacent tissue sections were probed for *WNT7B*, and quantification of RNA-ISH data are displayed in the table. Of nine primary tumors examined, five had >1% WNT5A expression in KRT+ cells (range 0.3%-42%) and seven had >1% WNT7B expression (range 0.5%-33.6%). Of 24 metastatic tumors examined, 16 had >1% WNT5A expression (range 0 to 50.5%) and 15 had >1% WNT7B expression (range 0 to 26%). (Right) Representative fluorescence micrographs of RNA in situ hybridization in prostate CTCs, probing for *WNT5A/7B* (yellow dots), and prostate CTC-specific markers (*EPCAM*, *KLK3*, *FOLH1*, *KRT8/18/19*) (red dots). DNA is stained with 4′,6′-diamidino-2-phenylindole (blue). Scale bar, 10 μm.

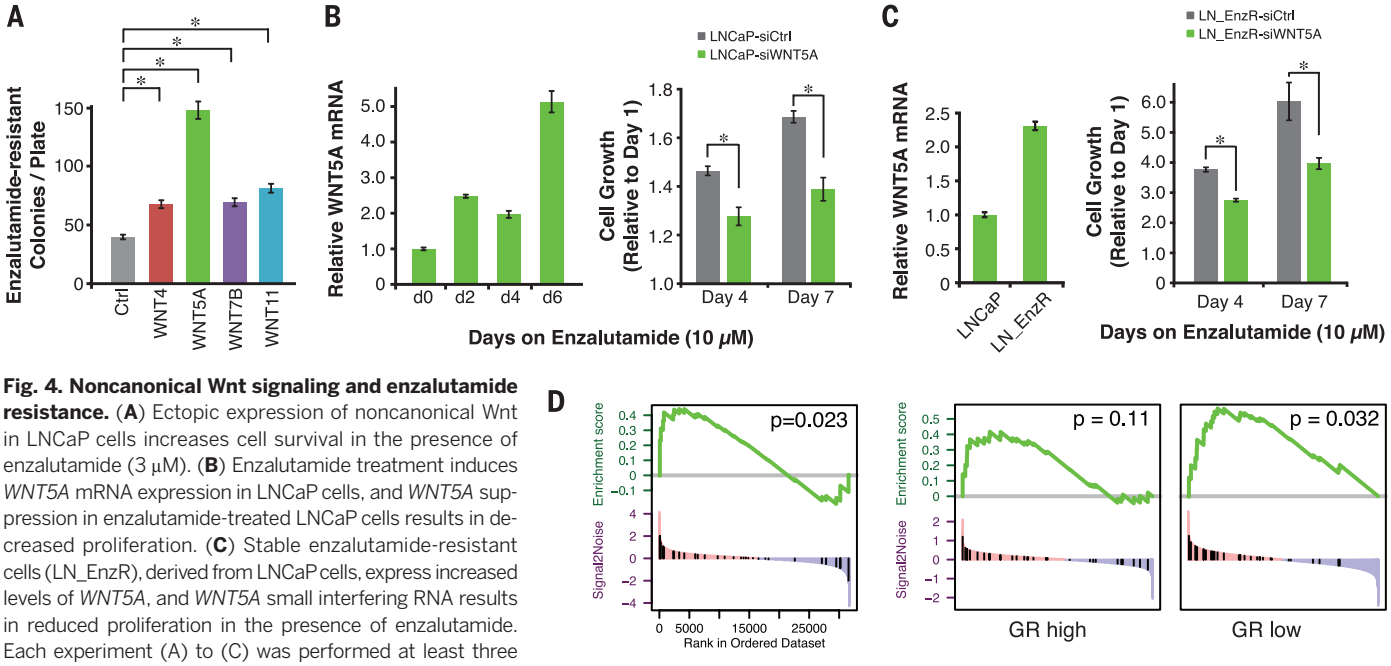


Fig. 4. Noncanonical Wnt signaling and enzalutamide resistance. (**A**) Ectopic expression of noncanonical Wnt in LNCaP cells increases cell survival in the presence of enzalutamide (3 μM). (**B**) Enzalutamide treatment induces *WNT5A* mRNA expression in LNCaP cells, and *WNT5A* suppression in enzalutamide-treated LNCaP cells results in decreased proliferation. (**C**) Stable enzalutamide-resistant cells (LN_EnzR), derived from LNCaP cells, express increased levels of *WNT5A*, and *WNT5A* small interfering RNA results in reduced proliferation in the presence of enzalutamide. Each experiment (A) to (C) was performed at least three times. Data are presented as means ± SD. (**D**) (Left) GSEA plot showing enrichment of noncanonical Wnt pathway in mouse xenografts derived from enzalutamide-resistant LREX′ cells (*29*) compared with control xenografts derived from LNCaP/AR cells (data from LREX′ and Con B entries, GEO GSE52169). (Right) GSEA plots showing noncanonical Wnt pathway enrichment in antiandrogen-resistant xenografts, when stratified by GR gene expression (data from Res and Con A entries, GEO GSE52169).

enzalutamide-progressing patients (group B) (P = 0.025), which was absent in CTCs with high *GR* expression (P = 0.34) (Fig. 3B and fig. S6D). Thus, these two AR-independent drug resistance pathways may predominate in different subsets of cancer cells.

Wnt proteins may be secreted by tumor cells as part of an autocrine loop, or they may be produced by surrounding stromal cells. We used RNA in situ hybridization (RNA-ISH) to identify the source of WNT production in tumor specimens and CTCs. Within primary untreated prostate cancers (n = 9), the noncanonical *WNT5A* and *WNT7B* mRNAs were present in a subset of tumor cells (8.9 and 11.6%, respectively), but both were rare in surrounding stromal cells (<0.2 and 0.5%, respectively) (Fig. 3C and fig. S6C). Metastatic tumor biopsies from patients with CRPC (n = 24) also had readily detectable *WNT5A* and *WNT7B* (8.0 and 6.1%, respectively) (Fig. 3C). Similarly, *WNT5A* or *WNT7B* mRNA was detected by RNA-ISH in a subset of CTCs from patients (n = 5) with CRPC (6 out of 180 CTCs; 3.3%) (Fig. 3C). Thus, a subset of prostate cancer cells express noncanonical Wnt ligands, which may provide survival signals in the context of AR inhibition.

To test whether activation of noncanonical Wnt signaling modulates enzalutamide sensitivity, we ectopically expressed the noncanonical ligands *WNT4*, *WNT5A*, *WNT7B*, or *WNT11* in LNCaP androgen-sensitive human prostate cancer cells, which express low endogenous levels (fig. S7, A and B). Survival of the AR-positive LNCaP cells in the presence of enzalutamide was enhanced by the noncanonical Wnt ligands, particularly *WNT5A* (Fig. 4A) ($P = 2.8 \times 10^{-5}$) (fig. S7C). Remarkably, endogenous *WNT5A* was acutely induced upon treatment with enzalutamide, suggestive of a feedback mechanism, and its depletion (knockdown) resulted in reduced cell proliferation (Fig. 4B and fig. S7D) ($P = 6.6 \times 10^{-4}$). We also generated stable enzalutamide-resistant LNCaP cells through prolonged in vitro selection (fig. S7E). These cells also exhibited increased expression of endogenous *WNT5A*, whose suppression reduced proliferation in enzalutamide-supplemented medium (Fig. 4C) (P = 0.005) (fig. S7F). Finally, we tested the contribution of noncanonical Wnt to antiandrogen resistance in an independent data set, interrogating the previously published mouse LNCaP xenograft model, in which aberrant activation of GR contributes to enzalutamide resistance (*29*). A significant association between enzalutamide resistance and noncanonical Wnt signaling was evident (P = 0.023), which again showed an inverse relation between *GR* expression and noncanonical Wnt signaling (P = 0.032 for *GR* low versus P = 0.11 for *GR* high) (Fig. 4D and fig. S8, A and B). This independent data set further validates the independent contributions of GR and noncanonical Wnt signaling to antiandrogen resistance.

In summary, by RNA profiling single prostate CTCs, we demonstrate their differences from primary tumors, as well as their heterogeneity within individual patients. The acquisition of AR-dependent and AR-independent alterations conferring resistance to antiandrogen therapies is also heterogeneous. Among *AR* alterations, more than half of all patients had multiple *AR* splice variants present within different CTCs and about 1 out of 6 of single cancer cells had simultaneous expression of several *AR* splice variants. Two AR-independent pathways, activation of GR and noncanonical Wnt signaling, coexist in different subsets of cells. Wnt signaling has been implicated in multiple cellular functions linked to prostate cancer progression (*4*, *25*–*28*), and noncanonical Wnt signaling may be targeted by suppression of its key downstream components, such as Rho kinase (*30*). Our study is limited by its retrospective nature and relatively small sample size (13 patients; average of six CTCs per patient), a consequence of the rarity of intact CTCs and inefficiencies inherent in manual single-cell micromanipulation techniques, obstacles that might be overcome with future improvements in CTC isolation and single-cell sequencing technologies. Nevertheless, the heterogeneity of CTCs in patients with CRPC stands in contrast to the striking homogeneity of AR signaling in single CTCs from untreated patients (*5*). Although these observations require validation in prospective trials, they point to complex and heterogeneous drug resistance mechanisms in advanced prostate cancer, which may affect therapeutic efficacy.

REFERENCES AND NOTES

1. A. Egan *et al.*, *Cancer Treat. Rev.* **40**, 426–433 (2014).
2. M. Gerlinger *et al.*, *N. Engl. J. Med.* **366**, 883–892 (2012).
3. H. I. Scher, M. J. Morris, S. Larson, G. Heller, *Nat. Rev. Clin. Oncol.* **10**, 225–234 (2013).
4. D. Robinson *et al.*, *Cell* **161**, 1215–1228 (2015).
5. D. T. Miyamoto *et al.*, *Cancer Discov* **2**, 995–1003 (2012).
6. D. Ramsköld *et al.*, *Nat. Biotechnol.* **30**, 777–782 (2012).
7. J. G. Lohr *et al.*, *Nat. Biotechnol.* **32**, 479–484 (2014).
8. N. Aceto *et al.*, *Cell* **158**, 1110–1122 (2014).
9. D. T. Ting *et al.*, *Cell Reports* **8**, 1905–1918 (2014).
10. E. Ozkumur *et al.*, *Sci. Transl. Med.* **5**, 179ra47 (2013).
11. Materials and methods are available as supporting material on *Science* Online.

12. B. Sharpe, M. Beresford, R. Bowen, J. Mitchard, A. D. Chalmers , *Stem Cell Rev.* **9**, 721–730 (2013).
13. J. Trepel, M. Mollapour, G. Giaccone, L. Neckers, *Nat. Rev. Cancer* **10**, 537–549 (2010).
14. T. Gutschner, M. Hämmerle, S. Diederichs, *J. Mol. Med. (Berlin)* **91**, 791–801 (2013).
15. C. F. Schaefer *et al.*, *Nucleic Acids Res.* **37** (Database), D674–D679 (2009).
16. B. Gottlieb, L. K. Beitel, A. Nadarajah, M. Paliouras, M. Trifiro, *Hum. Mutat.* **33**, 887–894 (2012).
17. C. Lu, J. Luo, *Transl. Androl. Urol.* **2**, 178–186 (2013).
18. E. S. Antonarakis *et al.*, *N. Engl. J. Med.* **371**, 1028–1038 (2014).
19. M. D. Balbas *et al.*, *eLife* **2**, e00499 (2013).
20. J. D. Joseph *et al.*, *Cancer Discov* **3**, 1020–1029 (2013).
21. S. Carreira *et al.*, *Sci. Transl. Med.* **6**, 254ra125 (2014).
22. J. Zhang, J. L. Manley, *Cancer Discov* **3**, 1228–1237 (2013).
23. L. Kurlender *et al.*, *Biochim. Biophys. Acta* **1755**, 1–14 (2005).
24. H. I. Scher *et al.*, *N. Engl. J. Med.* **367**, 1187–1197 (2012).
25. M. Katoh, M. Katoh, *Clin. Cancer Res.* **13**, 4042–4045 (2007).
26. S. Takahashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4938–4943 (2011).
27. D. Zheng *et al.*, *Mol. Cancer Res.* **11**, 482–493 (2013).
28. T. S. Gujral *et al.*, *Cell* **159**, 844–856 (2014).
29. V. K. Arora *et al.*, *Cell* **155**, 1309–1322 (2013).
30. N. Rath, M. F. Olson, *EMBO Rep.* **13**, 900–908 (2012).
31. F. Tang *et al.*, *Nat. Methods* **6**, 377–382 (2009).

ACKNOWLEDGMENTS

We thank C. Sawyers for helpful discussions; A. McGovern, E. Stadtmueller, and B. Abebe for clinical trial support; and L. Libby and L. Nieman for technical assistance. This work was supported by grants from the Prostate Cancer Foundation (D.A.H., S.M., M.T., M.R.S., and R.J.L.), Charles Evans Foundation (D.A.H.), Department of Defense (D.T.M., R.J.L., and D.T.T.), Stand Up to Cancer (D.A.H., M.T., S.M., and L.V.S.), Howard Hughes Medical Institute (D.A.H.), National Institute of Biomedical Imaging and Bioengineering (NIBIB), NIH, EB008047 (M.T.), NCI 2R01CA129933 (D.A.H.), National Cancer Institute, NCI, Federal Share Program and Income (S.M. and D.T.M.), Affymetrix, Inc. (D.T.T., K.A., and N.D.), Mazzone Program–Dana-Farber Harvard Cancer Center (D.T.M.), Burroughs Wellcome Fund (D.T.T.), and the Massachusetts General Hospital–Johnson & Johnson Center for Excellence in CTC Technologies (D.A.H., M.T., and S.M.). D.T.T. is a paid consultant for Affymetrix, Inc.; R.J.L. is a paid consultant for Janssen LLC. The Massachusetts General Hospital has filed for patent protection for the CTC-iChip technology. RNA-sequencing data have been deposited in GEO under accession number GSE67980.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1351/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S7
References (*32*–*37*)

10 March 2015; accepted 3 August 2015
10.1126/science.aab0917

SMALL PEPTIDES

Pri sORF peptides induce selective proteasome-mediated protein processing

J. Zanet,[1,2]* E. Benrabah,[1,2]* T. Li,[3] A. Pélissier-Monier,[1,2] H. Chanut-Delalande,[1,2] B. Ronsin,[1,2] H. J. Bellen,[3,4] F. Payre,[1,2]† S. Plaza[1,2]†

A wide variety of RNAs encode small open-reading-frame (smORF/sORF) peptides, but their functions are largely unknown. Here, we show that *Drosophila polished-rice* (*pri*) sORF peptides trigger proteasome-mediated protein processing, converting the Shavenbaby (Svb) transcription repressor into a shorter activator. A genome-wide RNA interference screen identifies an E2-E3 ubiquitin-conjugating complex, UbcD6-Ubr3, which targets Svb to the proteasome in a *pri*-dependent manner. Upon interaction with Ubr3, Pri peptides promote the binding of Ubr3 to Svb. Ubr3 can then ubiquitinate the Svb N terminus, which is degraded by the proteasome. The C-terminal domains protect Svb from complete degradation and ensure appropriate processing. Our data show that Pri peptides control selectivity of Ubr3 binding, which suggests that the family of sORF peptides may contain an extended repertoire of protein regulators.

Eukaryotic genomes encode many noncoding RNAs (ncRNAs) that lack the classical hallmarks of protein-coding genes. However, both ncRNAs and mRNAs often contain small open reading frames (sORFs), and there is growing evidence that they can produce peptides, from yeast (*1*) to plants (*2*, *3*) or humans (*4*, *5*). The *polished rice* or *tarsal-less* (*pri*) RNA contains four sORFs that encode highly related 11– to 32–amino acid peptides, required for embryonic development across insect species (*6*–*8*). In flies, *pri* is essential for the differentiation of epidermal outgrowths called trichomes (*7*, *8*). Trichome development is governed by the Shavenbaby (Svb) transcription factor (*9*–*11*); however, only in the presence of *pri* can Svb turn on the program of trichome development, i.e., activate expression of cellular effectors (*12*, *13*). Indeed, the Svb protein is translated as a large repressor, *pri* then induces truncation of its N-terminal region, which leads to a shorter activator (*12*). Thereby, *pri* defines the developmental timing of epidermal differentiation, in a direct response to systemic ecdysone hormonal signaling (*14*). Although we now have a clear framework for the developmental functions of *pri*, how these small peptides can trigger Svb processing is unknown.

To identify factors required for Svb processing in response to *pri*, we performed a genome-wide RNA interference (RNAi) screen in a cell line co-expressing green fluorescent protein (GFP)–tagged Svb and *pri* (Fig. 1A). We set up an automated assay quantifying Svb processing for each of the *Drosophila* genes, with an inhibitory score reflecting the proportion of cells unable to cleave off the Svb N terminus (see the supplementary materials). *pri* RNAi displayed the highest score, which validated our approach to identifying molecular players in Svb processing. Methods used to evaluate results from genome-wide screening all converged on a key role for the proteasome. For instance, COMPLEAT, a bioinformatic frame-

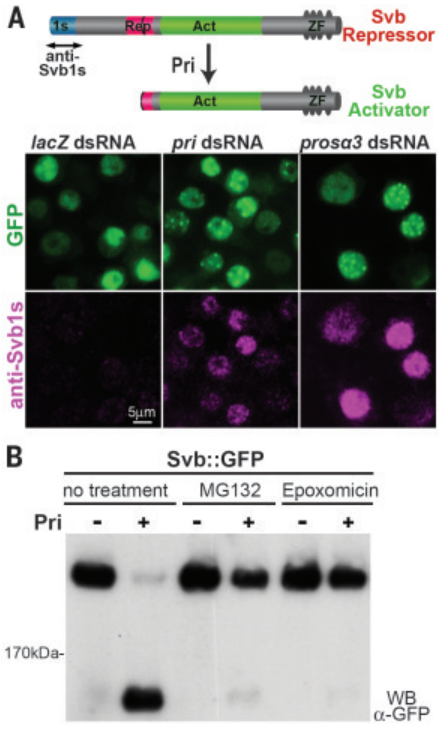


Fig. 1. Pri-dependent processing of Svb requires proteasome activity. (**A**) Drawing of Svb processing (antibody against Svb1s recognizes the repressor-specific N-terminal region) and snapshots from the screen illustrating the effect of double-stranded RNA against *lacZ* (negative control), *pri*, and *proteasomeα3* subunit (*prosα3*) on Svb::GFP processing. Cells were stained for Svb1s (purple) and GFP (green). (**B**) Western blot analysis of cells that express Svb::GFP, with or without *pri* and proteasome inhibitors (MG132, epoxomicin).

[1]Centre de Biologie du Développement, Université de Toulouse III—Paul Sabatier, Bâtiment 4R3, 118 route de Narbonne, F-31062 Toulouse, France. [2]CNRS, UMR5547, Centre de Biologie du Développement, F-31062 Toulouse, France. [3]Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA. [4]Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Neurological Research Institute, Baylor College of Medicine, Houston, TX 77030, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: francois.payre@univ-tlse3.fr (F.P.); serge.plaza@univ-tlse3.f (S.P.)**

work based on protein complex analysis (*15*), identified the proteasome in 66 out of the 71 top predictions (fig. S1A and table S1). A survey of individual proteasome subunits indicated that both the 20*S* catalytic core and the 19*S* regulatory particles are required for Svb processing (fig. S1B and table S2). Chemical proteasome inhibitors independently confirmed this conclusion, because they also prevented *pri*-induced Svb processing (Fig. 1B). These data thus provide compelling evidence that Svb processing results from a *pri*-dependent proteolysis by the proteasome.

To investigate how *pri* regulates proteolysis of Svb, we first identified the protein region(s) in Svb that are involved in *pri*-dependent processing. Systematic deletions demonstrated the importance of the Svb N terminus for *pri* response and restricted the minimal motif to the N-terminal 31 amino acids (fig. S2, A and B). Deletion of this motif within an otherwise full-length protein (Δ31) made Svb refractory to *pri* (fig. S2, A and B). Conversely, the Svb N terminus when fused to GFP (1s::GFP) was sufficient to transform this protein into a *pri* target and to make GFP sensitive to *pri*. Unlike Svb, however, 1s::GFP was completely degraded by the proteasome upon *pri* expression (Fig. 2A and fig. S2, C and D).

Recent studies have shown that structural features of proteins influence their degradation by the proteasome (*16*): Whereas unstructured substrates, such as intrinsically disordered regions, favor degradation (*17*), tightly folded domains can resist proteasome progression (*18*). Analysis of Svb sequences predicted intrinsically disordered features (Fig. 2A and fig. S3A) throughout its N-terminal moiety, which is degraded. By contrast, the proteasome-resistant C-terminal moiety comprises two folded regions: the transcriptional activation and zinc finger domains. Within the transcriptional activation region, we found that amino acids 532 to 701 protected Svb from complete degradation. Indeed, the C-terminally truncated mutants of 1 to 701 amino acids (and longer) were still processed, whereas mutants shortened by 1 to 532 amino acids (and shorter) were fully degraded (fig. S3B). We tested whether other folded domains would also protect Svb from complete degradation and found that attaching zinc fingers to short Svb mutants—otherwise degraded upon *pri* expression—was sufficient to restore processing (Fig. 2A). Likewise, the DNA binding domain of Gal4 protected against degradation (fig. S3, B and C), which indicated that even a heterologous protein domain with strong structure can protect Svb from full degradation in response to *pri*. Hence, distinct regions of Svb mediate its processing by the proteasome: the 31 N-terminal residues act as a *pri*-dependent degradation signal, or degron, and C-terminal domains act as stabilizing features that prevent complete degradation.

Proteins are targeted to the proteasome by the covalent attachment of ubiquitin to Lys residues (*19*). The Svb N terminus is highly conserved from insects to human (Fig. 2B); it comprises two invariant Lys residues (K3 and K8) and a third one at a less constrained position (K28 in *Drosophila*). We found that individual Lys substitutions had only a weak effect or no effect, whereas simultaneous mutation of all three Lys (3Kmut) abolished Svb processing (Fig. 2B). Furthermore, we detected strong *pri*-dependent ubiquitination of Svb when the proteasome was inhibited (Fig. 2C). By contrast, this was no longer seen in the 3Kmut variant, which demonstrated the key role of these three Lys in ubiquitin-dependent Svb processing.

Ubiquitin conjugation requires three enzymes (E1, E2, and E3); specificity is generally conferred by the E3 ubiquitin ligases that recognize and bind to substrates (*19*). A prominent hit from our RNAi screen was *Ubr3* (7 hits out of the top 15), which encodes an E3. Ranking all *Drosophila* ubiquitin enzymes by their inhibitory score confirmed that Ubr3 was the major E3 required for Svb processing and identified UbcD6 (Rad6) as its associated E2 (figs. S4 and S5 and table S3), consistent with evidence that human Ubr3 also forms a complex with UbcD6 (*20*). Like many

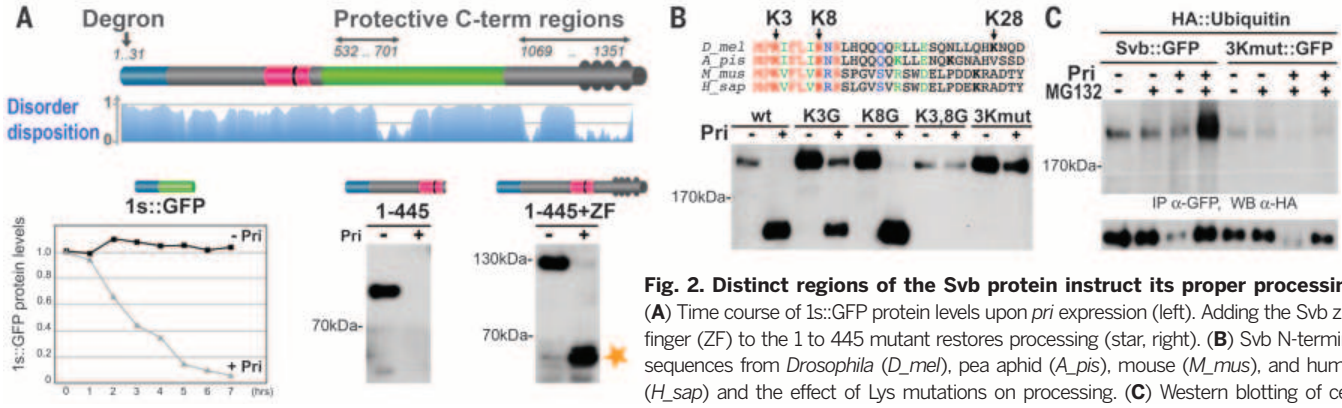


Fig. 2. Distinct regions of the Svb protein instruct its proper processing. (**A**) Time course of 1s::GFP protein levels upon *pri* expression (left). Adding the Svb zinc finger (ZF) to the 1 to 445 mutant restores processing (star, right). (**B**) Svb N-terminal sequences from *Drosophila* (*D_mel*), pea aphid (*A_pis*), mouse (*M_mus*), and human (*H_sap*) and the effect of Lys mutations on processing. (**C**) Western blotting of cells coexpressing hemagglutinin (HA)-tagged ubiquitin, Svb::GFP, or the 3Kmut variant, with or without *pri* and MG132.

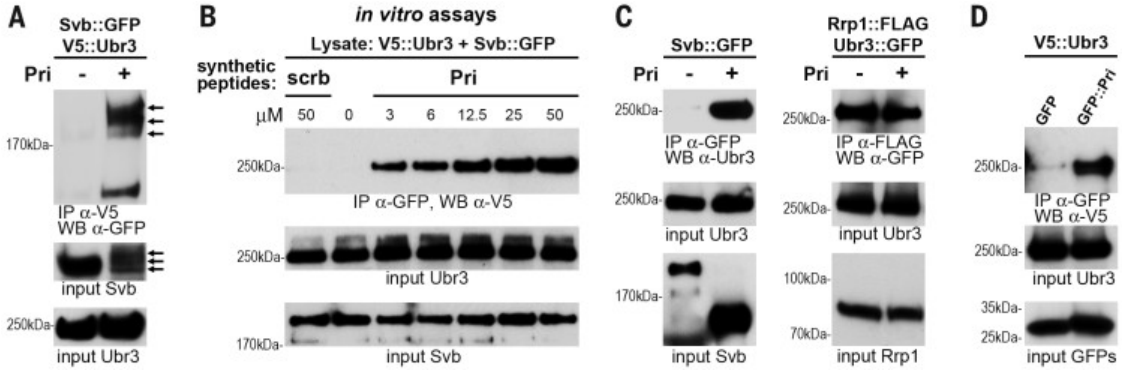


Fig. 3. Ubr3 mediates Svb ubiquitination in the response to Pri peptides. (**A**) Coimmunoprecipitation of Ubr3::V5 with Svb::GFP in the presence of MG132. Arrows indicate ubiquitinated Svb. (**B**) Protein extracts from cells coexpressing V5::Ubr3 and Svb::GFP, in the absence of *pri*, were incubated in vitro with a synthetic Pri peptide (MAAYLDPTGQY) (*31*) and immunoprecipitated using antibody against GFP. A scrambled peptide (MKTYPGALYDA, scrb) was used for control. (**C**) Coimmunoprecipitation of Ubr3 with Svb::GFP or with Rrp1::FLAG in the presence of MG132. (**D**) GFP and GFP::Pri were immunopurified and incubated with V5::Ubr3 protein extracts. Bound fractions were analyzed by antibody against V5 and Western blotting.

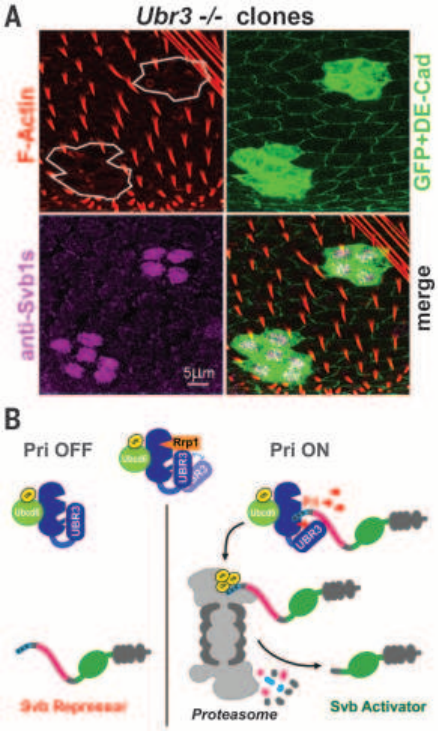


Fig. 4. *Ubr3* is required in vivo for Svb processing. (**A**) Clones of *Ubr3*-null epidermal cells in mosaic pupae (GFP-positive, green cytoplasm) do not form trichomes (F-actin, red) and retain unprocessed Svb repressor (Svb1s-specific antibody, purple). Cells contours are revealed by DE-cadherin (green). (**B**) Model of Svb processing in response to *pri*. After the binding of Pri peptides, Ubr3 becomes able to bind the Svb N terminus and, together with UbcD6, ubiquitinates three Lys residues. N-terminal unstructured regions of Svb are degraded by the proteasome, whereas C-terminal folded regions (green and gray ovals) protect from complete degradation and allow release of the truncated Svb activator. In contrast, Ubr3 binds to other substrates (e.g., Rrp1 and DIAP1), with or without Pri peptides.

proteasome factors, Ubr3 has a broad subcellular distribution in cytoplasm and nuclei, whereas Svb and UbcD6 are nuclear proteins (fig. S6). Svb processing still occurred normally when nuclear export was impaired (figs. S6 and S7 and table S4), which indicated that the proteolytic activation of Svb takes place within the nucleus.

Several additional lines of evidence support the conclusion that Ubr3 mediates the function of *pri* for Svb ubiquitination. First, Ubr3 coimmunoprecipitated with Svb in a *pri*-dependent manner and ubiquitinated Svb was found in a complex with Ubr3 upon proteasome inhibition (Fig. 3A). Second, the N terminus of Svb was sufficient for Ubr3 binding in response to *pri* (fig. S8). Note that a functional N-terminal degron in Svb was required for its interaction with Ubr3, because the ubiquitin-resistant 3Kmut variant no longer bound Ubr3. Third, in protein extracts from cells that do not express *pri*, addition of synthetic Pri peptide was sufficient to promote Ubr3-Svb interaction in vitro, in a dose-dependent manner (Fig. 3B). By contrast, a peptide of the same composition but in a "scrambled" sequence lacked activity.

Although critical for the binding of Ubr3 to the Svb N terminus, Pri peptides are, however, not indispensable for Ubr3 activity. We found that *pri* did not influence the binding of Ubr3 to Ape1 (Rrp1) (Fig. 3C), a factor involved in DNA repair and regulated by Ubr3-dependent proteasome degradation (*21*). Also, the interaction of Ubr3 with DIAP1, which inhibits apoptosis (*22*, *23*), occurred with or without *pri* (fig. S9). Moreover, we found that Pri peptides interacted with Ubr3, even in the absence of Svb (Fig. 3D and fig. S8). Finally, the isolated UBR-box of Ubr3 no longer required Pri peptides to bind Svb (fig. S10), which suggested that other Ubr3 motifs prevent Svb interaction in the absence of *pri*. We therefore conclude that Pri peptides directly regulate the selectivity of Ubr3 for binding to the Svb N terminus and, thereby, trigger Svb ubiquitination and processing by the proteasome.

We recently isolated a *Ubr3* loss-of-function allele (*24*) and assayed its phenotype in the differentiation of epidermal cells. As observed for *pri* mutants, embryos lacking *Ubr3* were unable to differentiate trichomes and to process Svb (fig. S11). Moreover, inactivation of either *UbcD6* or *Ubr3* prevented formation of adult trichomes in mosaic animals (Fig. 4A and fig. S12). When compared with their wild-type neighbors, *Ubr3*-null cells accumulated the repressor form of Svb, which demonstrated Ubr3's essential role for Svb processing in vivo.

Taken together, our data show that Pri peptides control the binding of the Ubr3 ubiquitin ligase to Svb and activate its processing by the proteasome (see Fig. 4B). In the absence of Pri, Ubr3 nonetheless recognizes other substrates (*21*–*23*), which shows that a main role for Pri peptides is to modify the binding selectivity of Ubr3. This could potentially be achieved through a conformational change in Ubr3 protein, as proposed for Ubr1 (*25*), that unmasked the recognition site for Svb upon Pri peptide binding to Ubr3.

Although recent work has uncovered thousands of novel sORF peptides (*1*–*5*), only a handful of their molecular targets have yet been identified. sORF peptides have recently been found to bind and regulate the Ca^{2+} uptake SERCA protein (*26*, *27*), the heterotrimeric guanine nucleotide–binding protein–coupled signaling APJ (Apelin) (*28*), and the DNA repair protein Ku (*29*). Protein-protein interactions often involve small protein regions, and artificial peptides that mimic these binding surfaces have been proven to be potent modulators of protein complexes (*30*). We propose that sORF-encoded peptides provide an unexplored reservoir of protein-binding interfaces, well suited to regulate the activity of a wide range of cellular factors.

REFERENCES AND NOTES

1. J. E. Smith *et al.*, *Cell Reports* **7**, 1858–1866 (2014).
2. K. Hanada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2395–2400 (2013).
3. D. Lauressergues *et al.*, *Nature* **520**, 90–93 (2015).
4. N. T. Ingolia *et al.*, *Cell Reports* **8**, 1365–1379 (2014).
5. J. Ma *et al.*, *J. Proteome Res.* **13**, 1757–1765 (2014).
6. J. Savard, H. Marques-Souza, M. Aranda, D. Tautz, *Cell* **126**, 559–569 (2006).
7. T. Kondo *et al.*, *Nat. Cell Biol.* **9**, 660–665 (2007).
8. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, J. P. Couso, *PLOS Biol.* **5**, e106 (2007).
9. F. Payre, A. Vincent, S. Carreno, *Nature* **400**, 271–275 (1999).
10. D. L. Stern, N. Frankel, *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20130028 (2013).
11. J. Crocker *et al.*, *Cell* **160**, 191–203 (2015).
12. T. Kondo *et al.*, *Science* **329**, 336–339 (2010).
13. D. Menoret *et al.*, *Genome Biol.* **14**, R86 (2013).
14. H. Chanut-Delalande *et al.*, *Nat. Cell Biol.* **16**, 1035–1044 (2014).
15. A. Vinayagam *et al.*, *Sci. Signal.* **6**, rs5 (2013).
16. S. Fishbain *et al.*, *Nat. Struct. Mol. Biol.* **22**, 214–221 (2015).
17. S. Prakash, L. Tian, K. S. Ratliff, R. E. Lehotzky, A. Matouschek, *Nat. Struct. Mol. Biol.* **11**, 830–837 (2004).
18. L. Tian, R. A. Holmgren, A. Matouschek, *Nat. Struct. Mol. Biol.* **12**, 1045–1053 (2005).
19. A. Varshavsky, *Annu. Rev. Biochem.* **81**, 167–176 (2012).
20. T. Tasaki *et al.*, *J. Biol. Chem.* **282**, 18510–18520 (2007).
21. C. Meisenberg *et al.*, *Nucleic Acids Res.* **40**, 701–711 (2012).
22. M. Ditzel *et al.*, *Mol. Cell* **32**, 540–553 (2008).
23. Q. Huang *et al.*, *Cell Death Differ.* **21**, 1961–1970 (2014).
24. S. Yamamoto *et al.*, *Cell* **159**, 200–214 (2014).
25. G. C. Turner, F. Du, A. Varshavsky, *Nature* **405**, 579–583 (2000).
26. E. G. Magny *et al.*, *Science* **341**, 1116–1120 (2013).
27. D. M. Anderson *et al.*, *Cell* **160**, 595–606 (2015).
28. A. Pauli *et al.*, *Science* **343**, 1248636 (2014).
29. S. A. Slavoff, J. Heo, B. A. Budnik, L. A. Hanakahi, A. Saghatelian, *J. Biol. Chem.* **289**, 10950–10957 (2014).
30. A. Patgiri, K. K. Yadav, P. S. Arora, D. Bar-Sagi, *Nat. Chem. Biol.* **7**, 585–587 (2011).
31. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

ACKNOWLEDGMENTS

We thank N. Perrimon, S. Mohr, O. Coux, P. Meier, N. Tapon, P. Demange, L. Twyffels, the *Drosophila* Screening Center, C. Polesello, M. Soulard, Y. Latapie, and P. Valenti for invaluable support. This work was supported by Agence Nationale de la Recherche (smORFpep and ChronoNet), Association pour la Recherche sur le Cancer, Fondation Recherche et Innovation Thérapeutique en Cancérologie, and Howard Hughes Medical Institute. The authors declare no conflict of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/349/6254/1356/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S4
References (*32*–*40*)

14 May 2015; accepted 10 August 2015
10.1126/science.aac5677

By **Michael P. Marshak**

My trek back to science

After I finished my Ph.D. studies at the Massachusetts Institute of Technology in 2012, my career was at a crossroads. I had gone to graduate school with the hope of pursuing a career in academia, but by the time I got my degree, that dream seemed far from reality. I was burned out, and I couldn't even bring myself to look for postdoctoral positions, let alone consider a lifetime in the lab. I also knew I couldn't possibly make an objective decision about what I should do for the rest of my life. I needed a break.

I am often drawn to extremes, so I thought: "What could possibly be further from a Ph.D. in chemistry than to thru-hike the Pacific Crest Trail (PCT), traversing the United States from Mexico to Canada?"

It was May and the sun was blazing at the California-Mexico border when I began my 2668-mile journey. Just days before I had been working in a chemistry lab, conducting research that I could take in seemingly infinite directions, but now there was just one: north. I started walking.

Over the next several weeks, my focus narrowed until my biggest concern was the number of miles to the next water source. I simply wanted to physically survive each day, which was a welcome change from the previous 5 years. In contrast with my experience in graduate school, progress on the trail was well defined and quantifiable: Each step brought me closer to Canada. Occupied with satisfying my immediate physical needs, I didn't spend much time consciously thinking about the future.

"I was burned out. ... I needed a break."

But after 16 weeks, 4 days, and 7 hours, I reached the Canadian border and my hiking adventure was over. Reality set in: I needed to return home and figure out what I was going to do next. I caught a bus from the trail's northern terminus at the Canadian border to Vancouver, then another bus to Seattle, Washington, then a flight back to Boston. I returned home 45 pounds under my typical healthy weight, as it had been difficult for me to consume all the calories I needed to hike nearly a marathon a day. While I recovered, I needed to eat every few hours and walk several miles each day to prevent leg spasms. It also took time to readjust emotionally to a life where there was no obvious, well-trodden path to follow.

I felt bored and detached after my trek, but this only made me more eager to start a new journey. Initially, I wasn't sure I was any closer to making a career decision than I had been before I started my hike, but after a few weeks of rest and recovery, it became clear that the weeks on the trail had refreshed me and renewed my love of science. I wanted to read journal articles to learn of new discoveries that had been reported while I was away. I was eager to embark on independent research. Most importantly, I missed the community of academics I had left behind.

So I decided I wanted to do a postdoc after all. I set up a meeting with my graduate school adviser, who then put me in contact with the lab head who would become my postdoctoral adviser. His research team was designing new organic flow battery chemistries, and it was a perfect fit for everyone: They got the synthetic chemist they needed, and I got a chance to work in an applied field with engineers, physical chemists, and theoreticians.

After hiking, I began to think much more about big-picture ideas than I had in graduate school. Although my Ph.D. said "Inorganic Chemistry," to the motley group of hikers I encountered on the trail I was simply a scientist. I took that identity to heart and came back to the lab with fewer inhibitions about learning new fields of research. I had also become more accepting of the messiness of cutting-edge science. Basic research can be slow and inefficient, and sometimes it's not clear if the effort will be worth the reward of discovery. But hiking the PCT helped me learn to celebrate little accomplishments, and to take each day and each experiment one step at a time—lessons that have served me well as I pursue my academic career path.

Michael P. Marshak is an assistant professor at the University of Colorado, Boulder, and a fellow at the Renewable and Sustainable Energy Institute. For more on life and careers, visit sciencecareers.org. Send your story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Mirus | The Transfection Experts

FREEDOM OF EXPRESSION

NEW! CHOgro™ Expression System

At Mirus Bio, we know it's all about expression. Introducing the new CHOgro™ Expression System, a transient transfection platform that finally gets high protein titers with robust cell growth in the most relevant CHO cells.

- **Efficient** – Enables high protein titers with simple workflow
- **Convenient** – Quick adaptation to CHO cell line lineages
- **Optimized** – High density growth with minimal clumping post-transfection
- **Worry-free** – No commercial license required; animal origin free

Available as a complete system or components sold separately.

Visit www.mirusbio.com/chogro to qualify for a CHOgro™ Expression System Trial Kit.

PASSION VISION INSPIRATION Celebrating 20 YEARS

mirusbio.com

Providing gene delivery expertise since 1995

"What I do with my Octet HTX time? Climb."

Shave weeks off your lead selection programs.

Broader antibody cross-competition ups your odds of finding the best candidates, but larger epitope binning studies take time. The Octet HTX system lets you use any binning assay format, any size matrix, start a run and get analyzed results the same day or the next day for larger studies. You can also combine multiple experiments into one dataset to easily visualize and cluster antibodies in similar bins or binding groups.

Lucy gets out of the lab more often now to climb. What will you do with your extra time?

fortéBIO®
A Division of ***Pall Life Sciences***

PALL Life Sciences

fortebio.com | 888-OCTET-75

Fast. Accurate. EASY.